DOCUMENT RESUME

ED 432 593 TM 029 970

AUTHOR	Davis, Rita, Ed.
TITLE	Proceedings of the Stake Symposium on Educational Evaluation (Champaign, Illinois, May 8-9, 1998).
PUB DATE	1998-05-00
NOTE	365p.; "With assistance from Marya Burke, Colleen Medley, and Theresa Souchet."
PUB TYPE	Collected Works - Proceedings (021)
EDRS PRICE	MF01/PC15 Plus Postage.
DESCRIPTORS	*Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Evaluation Problems; Higher Education; *Measurement Techniques; *Performance Based Assessment; Program Evaluation
IDENTIFIERS	*Stake (Robert E)

ABSTRACT

A symposium on educational evaluation was held to celebrate the career of Robert E. Stake. Contributions, which relate to many aspects of educational evaluation, include: (1) "The Issue of Advocacy in Evaluation" (Ernest House and Kenneth Howe); (2) "The Meaning of Bias" (Michael Scriven); (3) Commentary on Ernie House and Michael Scriven's Presentations (Lee Cronbach); (4) "Assessing, Evaluating, Knowing" (Linda Mabry); (5) "Balancing Philosophy and Practicality in Qualitative Evaluation" (Jennifer Greene); (6) "Naturalistic Generalizations as the Source of Investigative Insight" (Nick Smith); (7) "Naturalistic Generalizations: We Think What We Are" (Deborah Trumbull); (8) "What Is Really at Stake?" (Ulf Lundgren); (9) "Illogical Teaching" (Jim Raths); (10) "Love and Death and Responsive Evaluation" (Saville Kushner); (11) "Responsive Evaluation Amistad Style: Perspectives of One African-American Evaluator" (Stafford Hood); (12) "Who Knows? And Other Questions I Might Ask Bob Stake" (Sue Noffke); (13) "From Responsive to Collaborative Evaluation" (Rita O'Sullivan); (14) "Creating Evaluating Organizations" (James Sanders); (15) "'Give Me an Insight': Training and Reporting in Naturalistic Evaluation" (Helen Simons); (16) "Possibilities for Cultivating Evaluative Intelligence" (Lou Rubin); (17) "Evaluation Is Not Evaluation Is Not Evaluation" (Norm Stenzel); (18) "Setting Performance Standards for National Board Assessments: A Reprise on Research and Development" (Dick Jaeger); (19) "The Legacy of Centers" (Tom Fox); (20) "Case Study: The Importance of Multiple Takes" (Jacquie Hill); (21) "Robert Stake and Our Business of Evaluation" (Katherine Ryan and John Ory); (22) "Social Work Evaluation in Sweden and Robert Stake" (Haluk Soydan); (23) "Case Study Approach in the Negotiating Evaluation Model" (Maria Saez Brezmas and Antonio Carretero); (24) "Excerpts from an Evaluation of Kenwood Elementary School's Year-Round Program" (Delwyn Harnisch, Philip Zodhiates, and Najmuddin Shaik); (25) "A Study of an Empowered School: An Investigation of the Development and the Effect of a Teacher Empowerment Process" (Carmen Palmer); (26) "Effects of a Museum-School Collaborative on Seventh Grade Students of an Urban Public Elementary School" (Mary Ann Ludwig); (27) "Two Faces of Urban High School Students: Characteristics of Dropouts and Persisters" (Lois Gueno); (28) "Ghosts and Reminiscences: My Last Day on Earth as a 'Quantoid'" (Gene Glass); (29) "Bob Stake Meets Mr. Rogers" (David Balk); (30) "Tom and Bob: CIRCE '64 to '67: Evaluation Sweetwater on the Illinois Plains: Portrait of an Education: A Responsive Reflection: Five Colons in Search of a Paper" (Thomas O. Maguire); (31) "Two Measurement Guys

Gone Wrong or Fumbling and Stumbling toward a Paradigm" (Lou Smith); (32) "35 Years Goes Fast When You're Having Fun" (Les McLean); (33) "Soy(a) Bean Futures Near the Arctic Circle (or How Green Was Bob's Volvo?)" (David Hamilton); (34) "High Expectations at CIRCE: Bob as Mentor" (Theresa Souchet, Marya Burke, Chris Migotsky, Rita Davis, Edith Cisneros-Cohernour, and Mindy Miron Basi); and (35) "A Brazilian's Stakian Journey" (Iduina Chaves). Remarks about the career of Robert Stake from six other conference participants are included, along with a response by Robert Stake. Most papers contain references. (SLD)

****
*   Reproductions supplied by EDRS are the best that can be made   *
*   from the original document.   *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Rita Davis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 432 593

Proceedings of the

Stake Symposium on Educational Evaluation

The Symposium was held
May 8 and May 9, 1998
at the University of Illinois

Rita Davis, Editor,
with assistance from
Marya Burke, Colleen Medley, and Theresa Souchet

TM029970

Proceedings of the

Stake Symposium on Educational Evaluation

The Symposium was held
May 8 and May 9, 1998
at the University of Illinois

Rita Davis, Editor,
with assistance from
Marya Burke, Colleen Medley, and Theresa Souchet

Table of Contents

Page

Lizanne DeStefano, *Forward* i

Bob Stake, *Acknowledging* iii

Symposium Program v

List of Participants xiii

Introduction by Lee Cronbach 1

Ernest House and Kenneth Howe, *The Issue of Advocacy in Evaluation* 5

Michael Scriven, *The Meaning of Bias* 13

Lee Cronbach Commentary on House and Scriven 25

Linda Mabry, *Assessing, Evaluating, Knowing* 31

Jennifer Greene, *Balancing Philosophy and Practicality in Qualitative Evaluation* 35

Nick Smith, *Naturalistic Generalizations as the Source of Investigative Insight* 51

Deborah Trumbull, *Naturalistic Generalizations: We Think What We Are* 63

Ulf Lundgren, *What is Really at Stake?* 75

Jim Raths, *Illogical Teaching* 83

Saville Kushner, *Love and Death and Responsive Evaluation* 91

Stafford Hood, *Responsive Evaluation Amistad Style: Perspectives of One African-American Evaluator* 101

Sue Noffke, *Who Knows? And Other Questions I might ask Bob Stake* 113

Rita O'Sullivan, *From Responsive to Collaborative Evaluation* 121

James Sanders, *Creating Evaluating Organizations* 137

Helen Simons, *"Give Me An Insight": Training and Reporting in Naturalistic Evaluation* 141

Lou Rubin, *Possibilities For Cultivating Evaluative Intelligence* 155

Norm Stenzel, *Evaluation is not Evaluation is not Evaluation* 161

Dick Jaeger, *Setting Performance Standards for National Board Assessments: A Reprise on Research and Development* 173

Tom Fox, *The Legacy of Centers* 179

Jacquie Hill, *Case Study: The Importance of Multiple Takes* 189

Katherine Ryan and John Ory, *Robert Stake and Our Business of Evaluation* 201

Haluk Soydan, *Social Work Evaluation in Sweden and Robert Stake* 215

María Sáez Brezmas and Antonio Carretero, *Case Study Approach in the Negotiating Evaluation Model* 225

Delwyn Harnisch, Philip Zodhiates and Najmuddin Shaik, *Excerpts from An Evaluation of Kenwood Elementary School's Year-Round Program* 239

Carmen Palmer, *A Study of an Empowered School: An Investigation of the Development and the Effect of a Teacher Empowerment Process* 253

Mary Ann Ludwig, *Effects of a Museum-School Collaborative on Seventh Grade Students of an Urban Public Elementary School* 259

Lois Gueno, *Two Faces of Urban High School Students: Characteristics of Dropouts and Persisters* 267

And the more personal . . .

Gene Glass, *Ghosts and Reminiscences: My Last Day on Earth as a "Quantoid"* 277

David Balk, *Bob Stake Meets Mister Rogers* 291

Tom Maguire, *Tom and Bob: CIRCE '64 to '67: Evaluation Sweetwater on the Illinois Plains: Portrait of an Education: A Responsive Reflection: Five Colons in Search of a Paper* 303

Lou Smith, *Two Measurement Guys Gone Wrong or Fumbling and Stumbling Toward a Paradigm* 313

Les McLean, *35 Years Goes Fast When You're Having Fun* 319

David Hamilton, *Soy(a) Bean Futures near the Arctic Circle (or How Green was Bob's Volvo?)* 323

Theresa Souchet, Marya Burke, Chris Migotsky, Rita Davis, Edith Cisneros-Cohernour, and Mindy Miron Basi, *High Expectations at CIRCE: Bob as Mentor* 327

Iduina Chaves, *A Brazilian's Stakian Journey* 331

Mildred Griggs, *Remarks* 345

Jeff Stake, *Remarks* 347

Clem Adelman, *Remarks* 349

Dan Alpert, *Remarks* 351

Barry MacDonald, *Remarks* 353

Terry Denny, *Selected Memories of Robert Stake* 355

Intro by Elliot Eisner, *It's all about Bob Stake* 361

Bob Stake, *Hoax?* 363

Foreword

Lizanne DeStefano
University of Illinois

One May 8th and 9th, 1998 a group of scholars, students, family members and friends gathered on the campus of the University of Illinois to celebrate the career of Robert Earl Stake. Bob and I were amazed at their number. In our initial planning, we anticipated 35 out of town guests and perhaps the same number of local participants. When the day came, more than 250 people joined in. We filled the meeting rooms beyond capacity, taxed the caterer's good humor, and had to rent a bus to take everyone to dinner where we commandeered the entire restaurant. I have never been a part of anything like it.

Now, six months later, as I reflect on those two days and the months of planning that preceded them, my strongest impression is of the unique combination of personal and professional concern that permeated the event. People were motivated to travel long distances, make presentations and write papers because they wanted to acknowledge Bob as a major figure in the field of evaluation *and* as a significant influence in their lives. As I chatted with folks on the phone or over e-mail in the weeks before the conference I cannot tell you how many "Bob" stories I heard. The symposia presentations and formal remarks were rife with them. In these stories Bob's role ranged from matchmaker to critic, but time and time again, his friends, family, and colleagues told of how Bob's wit, cynicism, critical eye and unique perspective had changed the way they thought about something. Quite remarkable, I think.

From the beginning, we had intended for the symposium to result in a publication. When it was over, I lost some heart for that task. I felt at the time that a print volume can in no way capture what went on during those two days. Now the volume exists. It does not recreate the physical thrill of seeing the icons of our field lunching, laughing, talking about old times and thinking about the next generation of educational inquiry. It doesn't convey the poignancy of that moment when several generations of Bob's students reflected on his mentorship. It lacks the energy and good will that surrounded us during those two days. You had to be there. It is as simple as that.

The proceedings does give us a record of the fine thinking, care, and effort that folks brought to the Stake Symposium. It is a wonderful reflection on Bob's career and how his presence has influenced persons from all over the world in so many different ways. For those of us who participated, it sparks memories of those special moments throughout the conference and the specialness of being there.

Acknowledging

Bob Stake

One of our former Illinois colleagues, Dave Nyberg, wrote a fine little book, *The Varnished Truth*, not a bad theme for reading the pages ahead. Some will find more varnish than truth. Dave's point was that varnishing is an essential part of our humanity and culture. We may be our best selves at graduations, weddings and funerals. We were our best selves at this Symposium.

As Lizanne just said, you had to be there. It was a perfect ten. And largely because she put it together, thoughtfully, ingeniously, generously. It had Mildred's blessing and backing from the Jack Easley Endowment and the Daniel A. Alpert Fund. They had good help, to be sure: Elizabeth Easley, Karen Andrews, Beena Choksi, Connie Dorsett, Trudy Morritz, Diane Erdman-Hamer, Rita Davis, Theresa Souchet, Marya Burke, Edith Cisneros-Cohernour. One of the finest ever quilted. My appreciations of the occasion are spelled out further in the final piece in these proceedings.

I found it delicious to be celebrated. Reality banished for the day. One thing wrong, though: there was too little time for personal talk and deep reflection. Too little acknowledging of credit due. We in educational research, certainly program evaluation, have a lot of trouble with attribution. So many causes. So many connections. So many needed to be acknowledged, and not just on that occasion, but throughout this long career, so many that never got the credit due.

It is said that insanity is hereditary, you get it from your kids. I think the same is true of understanding. If you've got any, you surely get some it from your kids. And it's matrimonial. Jeff, in his remarks toward the end of this volume, rightly recognized that much of what I am and have been, I got from Bernadine.

So many to whom I owe so much. Especially intellectually. How thin the line between plagiarism and insight. We teach our students to think what we think more than to think what they think. My thoughts are my teachers' thoughts. Did I really ever have a thought of my own?

And there so many were, gathered at my Symposium, my teachers, elder and younger. I regularly thought of Ernie, Terry, Stephen, Linda, and others as youngsters, only slowly realizing they had outreached me, had shaped the thoughts I thought I was mentoring for them. Gene, I knew right away.

And so many who couldn't come, so many who poured a stream of their lives into me, especially Tom Hastings. And Jack Easley. And Arden Grotelueschen, Richard Madden, Warren Findley, Walt Sehnert, Chris Buethe, Carmilva Flores, Barry McGaw, David Metzer, Lydia Cochran, Chuck Neidt, Jean Stutt, Dale Bainbridge, Burt Evans, Laury Gulick, Jerry Cote, Ed Kelly, Sigbrit Franke-Wikberg, Erik Wallin, Wayne Welch, Mary Lee Smith, Buddy Peshkin, Kip Anastasiou, Larry Metcalf, Ron Palosaari, Harold Gulliksen, Warren Baller, Mamie Hickey, Jo Merrick, Jack Larson, Bill Surman, Marianne Amarel, Hal Taylor, Peter Taylor, Jennie Fleagle, Deborah Laughton, Tina Ekstrom, Mary Jean Davis, Paul Barton, Henry Kaiser, Bob Kalisch, Della Lewis, Christina Carvajal, Urban Dahllöf, Bob Long, Helen Rose, Dick Spencer, Ruth Dunham, Doug Sjogren, Peter Fensham, David Pearson, Carl Helm, Sam Webb, Hank Slotnick, Elmer Sprague, Ron Holt, Carol Wintermute, Royce Sadler, Jerry Hausman, Brent Wilson, Lloyd Teale, Mel Hesser, Eric Joselyn, Fannie Bates, Steph Simpson, Edna Kuster, Merl Malehorn, Jack Morrison, Ernie Olson, Decker Walker, Rob Walker, Fred Kling, Giordana Rabitti, Alan Lemke and Randy Lemke, Phil Sorensen, Kjell Hërnqvist, Harry Broudy, Ledyard Tucker, Jim Popham, Chuck Caruson, Gerry Gage, Alan Purves. And that's not the half of it. And especially, Tom Hastings.

But varnish and attribution notwithstanding, we're planning to get the whole group together in May, 2027. Y'all come.

Program

Stake Symposium on Educational Evaluation

Friday, May 8, 1998, Levis FacultyCenter

8:00-9:00 a.m. Registration and Coffee, Fourth Floor

9:00-10:00 a.m. Music Room

Chaired by **Oli Proppé**, Iceland Inst. of Educ., and **Penha Tres**, University of California, Irvine.

Bob Stake: *Welcome.*

Rita O'Sullivan, U. of NC, Greensboro: *From Responsive to Collaborative Evaluation.*

Jennifer Greene, Cornell University: *Balancing Philosophy and Practicality in Qualitative Evaluation.*

10:00-12:00 a.m. The following two sessions will be repeated both at 10 and 11 am

Room 401

Chaired by **David Hamilton**, University of Umeå, and **Henriette Heimgaertner**, Van Leer Fdn.

Deborah Trumbull, Cornell University: *Naturalistic Generalizations: We Are What We Think.*

Nick Smith, Syracuse University: *Naturalistic Generalizations as the Source of Investigative Insight.*

11:00-12:00 p.m. Music Room

Chaired by **Lawrence Ingvarson**, Monash U., and **David Pearson**, Michigan State University.

Dick Jaeger, U. of North Carolina, Greensboro: *What Cognitive and Social Psychology Imply about Setting Performance Standards.*

12:00-1:00 p.m. **Lunch (not arranged)**

1:00-2:00 p.m. Room 401

Chaired by **Claryce Evans**, Harvard University, and **David Jenness**, Valley Research, Santa Fe.

Jacquie Hill, University of Illinois:
Case Study: The Importance of Multiple "Takes."

María Sáez, University of Valladolid:
Case Study Approach in the Negotiating Evaluation Model.

1:00-2:00 p.m. Room 405 - 406

Chaired by **Jeri Nowakowski**, NCREL, and **Bill Foster**, National Center on Substance Abuse and Addiction at Columbia University.

James Sanders, Western Michigan University:
Creating Evaluating Organizations.

Tom Fox, National-Louis University:
The Legacy of Centers.

Music Room

Chaired by **Pat Templin**, Right Associates, Cupertino, CA. and **Kathryn Sloane**, University of Illinois.

Saville Kushner, University of East Anglia:
Love and Death and Responsive Evaluation.

Jim Pearsol, Ohio State University:
Responsive Evaluation as a Tool for Continuous Quality Improvement in the Public Sector.

2:00-3:00 p.m. Room 401

Chaired by **Nick Smith**, Syracuse University
and **Jennifer Greene**, Cornell University.

A discussion initiated by **Helen Simons**, Univ of Southhampton:

Insight: How to Achieve it, Especially in Case Study and Collaborative Evaluation.

Room 405 - 406

Chaired by **Nigel Norris**, University of East Anglia.

Lou Rubin, University of Illinois:

Cultivating Evaluative Intelligence.

Norm Stenzel, University of Illinois:

Evaluation is not evaluation is not evaluation.

Music Room

Chaired by **Fred Rodgers**, University of Illinois, and
Kristin Powell, Chicago Teachers Academy.

Mary Ann Ludwig, Chicago Public Schools:

Effects of a Museum-School Collaborative on Seventh Grade Students of an Urban Public Elementary School.

Lois Gueno, Chicago Public Schools:

Two Faces of Urban High School Students: Characteristics of Drop Outs and Persisters.

Carmen Palmer, Chicago Public Schools:

An Empowered School: An Investigation of the Development and the Effect of a Teacher Empowerment Process.

3:00-4:00 p.m. Room 401

Chaired by **Renée Clift,** Univ of Illinois and **John McLure,** Univ of Iowa.

Del Harnisch, **Philip Zodhiates** and **Naj Shaik,** U of I:

Evaluating Year Round Education Programs.

Philip Holmes-Smith, Victoria Dept of Educ:

Evaluating School Performance: Accountability and School Improvement.

Room 405 - 406

Chaired by **Ulf Lundgren**, Skolvorket, Stockholm, and **Barry MacDonald**, University of East Anglia.

Haluk Soydan, Swedish Board of Health and Welfare:

Evaluation and Social Work in Sweden.

Iduina Chaves, Fluminense Federal Univ.:

A Brazilian's Stakian Journey.

Music Room

Chaired by **Ken Komoski**, EPIE and **Dennis Gooler**, NCREL.

Chip Bruce, Univ of Illinois:

Evaluating Information Technologies.

David Balk, Oklahoma State University:

Bob Stake Meets Mister Rogers.

4:00-5:00 p.m. Music Room

Chaired by **Liora Bresler**, University of Illinois and **Gary Joselyn**, University of Minnesota.

Katherine Ryan and **John Ory**, U. of Illinois:

Robert Stake and the Business of Evaluation.

Stafford Hood, Arizona State University:

Responsive Evaluation Amistad Style: Perspectives of One African-American Evaluator.

5:00-7:00 p.m. **Reception**, Krannert Center for the Performing Arts, 500 S. Goodwin, Urbana

Mike Atkin, *Master of Ceremonies.*

Carmilva Flores, Chris Migotsky, Theresa Souchet, Rita Davis, Marya Burke, Edith Cisneros-Cohernour, and **Mindy Basi,** *High Expectations.*

And other words from **Mildred Griggs, Jeff Stake, Clem Adelman, Dan Alpert, Barry MacDonald, Madeleine Grumet, Terry Denny.**

Music by **Tim Green** and **Gary Cziko,** U of I,
Clem Adelman, Trondheim U.

7:30 p.m. Dinner, Shurts House Inn; see **Elizabeth Easley** about reservations, ride.

Saturday, May 9, 1998, Room 407, Levis Faculty Center

8:00-9:00 a.m.	Registration and Coffee
9:00-10:15 a.m.	Room 407

Opening Session
Ernest House, University of Colorado: *Values.*
Michael Scriven, Claremont University: *Bias.*
Introduction and commentary by **Lee Cronbach**, Stanford U.

10:15-10:30 a.m. **Break**

10:30-12:00 p.m. Room 407

Panel on: *Assessing, Evaluating, Knowing.*
Jim Raths, University of Delaware.
David Hamilton, University of Umeå.
Sue Noffke, University of Illinois.
Gene Glass, Arizona State University.
Moderated by **Linda Mabry**, Indiana University.

12:00-1:00 p.m. **Lunch**, Levis Faculty Center, Second Floor

1:00-2:30 p.m. Room 407

Presentations introduced by **Lizanne DeStefano**, U of I.

Tom Maguire, University of Alberta:
Thoughts of Tom.

Les McLean, University of Toronto:
Thirty-Five Years Goes Fast When You're Having Fun.

Lou Smith, Washington U. of St. Louis:
Two Measurement Guys Gone Wrong; Fumbling and Stumbling Toward a Paradigm.

Ulf Lundgren, Skolverket, Stockholm:
What is Really at Stake?

2:30-2:45 p.m. **Break**

2:45-3:30 p.m. Room 407

Closing Session

Introduction by **Elliot Eisner**, Stanford University.

Bob Stake, University of Illinois: *Hoax?*

3:30 p.m. **Reception**, Levis Faculty Center Reading Room

Planning: Lizanne DeStefano, Karen Andrews, Liora Bresler, Marya Burke, Beena Choksi, Rita Davis, Connie Dorsett, Elizabeth Easley, Diane Erdman-Hamer, Trudy Morritz, Terry Souchet.

Appreciation: This Symposium was made possible by support from the Bureau of Educational Research, the Jack Easley Endowment, and the Daniel A. Alpert Fund.

List of Participants

Terry Ackerman, *University of Illinois*
Clem Adelman, *Trondheim University*
Henry Akplu, *University of Illinois*
Daniel Alpert, *University of Illinois*
Tom Anderson, *University of Illinois*
Jenny Anderson, *Martinsville, Indiana*
Karen Andrews, *University of Illinois*
Dolores Appl, *University of Illinois*
Sheila Arens, *Indiana University*
Ann Atkin, *Palo Alto, California*
Mike Atkin, *Stanford University*
David Balk, *Oklahoma State University*
Mary Ann Balk, *Stillwater, Oklahoma*
Mindy Basi, *University of Illinois*
Alexis Benson, *University of Illinois*
Kenneth Benson, *University of Illinois*
Lanny Beyer, *Indiana University*
Eunice Boardman, *University of Illinois*
Ray Boehmer, *University of Illinois*
Larry Braskamp, *Loyola University (Chicago)*
Liora Bresler, *University of Illinois*
David Brevig, *San Clemente, California*
Christie Brinkley, *Champaign, Illinois*
Claire Brown, *Wadlamah Island, South Carolina*
David Brown, *University of Illinois*
Chip Bruce, *University of Illinois*
Susan Bruce, *University of Illinois*
Deborah Bruns, *University of Illinois*
Jim Buell, *University of Illinois*
Cheryl Bullock, *University of Illinois*
Mary Anne Bunda, *Western Michigan University*
Nick Burbules, *University of Illinois*
Chris Burke, *University of Illinois*
Marya Burke, *University of Illinois*
Donald Burkholder, *University of Illinois*
Jean Burkholder, *Urbana, Illinois*
Kristine Burnaska, *University of Illinois*
Frederick Burrack, *University of Illinois*
Georganne Burton, *Urbana, Illinois*
Vernon Burton, *University of Illinois*
Tone Carlsten, *Oslo, Norway*

Merrill Chandler, *University of Illinois*
Iduina Chaves, *Rio de Janeiro*
Roger Chesswas, *NCERL*
Beena Choksi, *University of Illinois*
Edith Cisneros-Cohernour, *University of Illinois*
Renee Clift, *University of Illinois*
Richard Colwell, *Boston, Massachusetts*
Rob Corso, *University of Illinois*
William Cresswell, *Urbana, Illinois*
Lee Cronbach, *Stanford University*
Larry Cross, *University of Illinois*
Gary Cziko, *University of Illinois*
Melissa Dark, *Purdue University*
Rita Davis, *University of Illinois*
Judy Dawson, *Illinois State Board of Education*
Beth Dawson, *Southern Illinois University*
Jo Ann Day, *Decatur, Illinois*
Connie Kubo-Della-Piana, *University of Texas at El Paso*
Gabe Della-Piana, *University of Texas at El Paso*
Terry Denny, *Tucson, Arizonia*
Phyllis Denny, *Tucson, Arizonia*
Wendy Denny, *Saratoga Springs, New York*
Norman Denzin, *University of Illinois*
Gary Depaul, *University of Illinois*
Lizanne DeStefano, *University of Illinois*
Skylar DeWitt, *Saratoga Springs, New York*
Deborah Dillon, *University of Illinois*
Elizabeth Easley, *Champaign, Illinois*
Elliot Eisner, *Stanford University*
Nancy Ellis, *University of New Hampshire*
Terry Elofson, *NWREL*
Dick Elsholz, *Bloomfield, Michigan*
Dorothy Espelage, *University of Illinois*
Claryce Evans, *Harvard University*
Walter Feinberg, *University of Illinois*
Rhoda Feldman, *Chicago, Illinois*
Belden Fields, *University of Illinois*
Janetta Fleming, *University of Illinois*
Bill Foster, *CASA, NYC*
Susan Fowler, *University of Illinois*
Tom Fox, *National Louis University*
Colleen Frost, *Brandon, Mississippi*
Janet Gaffney, *University of Illinois*
Melissa George, *Purdue University*
Mark Gierl, *University of Alberta*

Deb Gilman, *University of Illinois*
Gene Gilmore, *Urbana, Illinois*
Virginia Gilmore, *Urbana, Illinois*
Craig Gjerde, *University of Wisconsin*
Gene Glass, *Arizonia SE University*
Sandy Glass, *Arizonia SE University*
Corrine Glesne, *University of Utah*
David Goodwin, *St. Louis, Illinois*
Dennis Gooler, *NCREL*
Thomas Grayson, *University of Illinois*
Jennifer Greene, *Cornell University*
Tim Green, *Urbana, Illinois*
Mildred Griggs, *University of Illinois*
Lois Grimason-Hilton, *Champaign, Illinois*
Beth Grosshandler, *University of Illinois*
Dean Grosshandler, *University of Illinois*
Madeleine Grumet, *University of North Carolina*
Lois Gueno, *Chicago, Illinois*
Mel Hall, *North Arizona University*
Jim Halle, *University of Illinois*
David Hamilton, *Umeå University*
Jill Hamm, *University of Illinois*
Del Harnisch, *University of Illinois*
Pat Harnisch, *Urbana Public School District*
Doris Harvey, *Urbana, Illinois*
Dick Harvey, *Urbana, Illinois*
Eve Harwood, *University of Illinois*
Mark Hawkes, *NCREL*
Laird Heal, *University of Illinois*
Dan Heck, *University of Illinois*
Henriette Heimgaertner, *Bernard van Leer Foundation*
Nancy Hertzog, *University of Illinois*
Jacquetta Hill, *University of Illinois*
Harry Hilton, *University of Illinois*
Michelle Hinn, *University of Illinois*
Wells Hively, *St. Louis*
Gordon Hoke, *University of Illinois*
Philip Holmes-Smith, *Melbourne*
Stafford Hood, *Arizona State University*
Marty Hotvedt, *Allentown, Pennsylvania*
Ernest House, *University of Colorado*
Lloyd Humphreys, *Whiteheath, Illinois*
Lawrence Ingvarson, *Monash University*
Kayleen Irizarry, *Chicago, Illinois*
Richard Jaeger, *University of North Carolina, Greensboro*

Golie Jansen, *Cheney, Washington*
Maceo Jean-Pierre, *Albuqueque, New Mexico*
David Jenness, *Santa Fe, New Mexico*
Ken Jerich, *Illinois State University*
Edna Johnson, *University of Illinois*
Trav Johnson, *University of Illinois*
Ben Joselyn, *Urbana, Illinois*
Bernadine Joselyn, *Moscow*
Eleanor Joselyn, *Urbana, Illinois*
Gary Joselyn, *University of Minnesota*
Mark Joselyn, *University of Illinois*
Sam Joselyn, *Urbana, Illinois*
Yleen Joselyn, *Crystal, Minnesota*
Lilian Katz, *University of Illinois*
Tom Kerins, *Springfield, Illinois*
Susan Klonsky, *University of Illinois at Chicago*
Kim Knutson, *University of Illinois at Chicago*
Paula Kohler, *University of Illinois*
Ken Komoski, *EPIE*
Barbara Kremer, *Northwestern University*
Amy Kummerow, *Urbana, Illinois*
Fred Kummerow, *University of Illinois*
Saville Kushner, *University of East Anglia*
Ira Langston, *University of Illinois*
Pei-Hsuan Lin, *University of Illinois*
Sharon Litchfield, *University of Illinois*
Jane Loeb, *University of Illinois*
Mary Ann Ludwig, *Chicago, Illinois*
Ulf Lundgren, *Stockholm*
Linda Mabry, *Indiana University*
Barry MacDonald, *University of East Anglia*
Wendy Madden, *University of Illinois*
Martin Maehr, *University of Michigan*
Jane Maehr, *High Scope*
Micheline Magnotta, *University of Illinois*
Tom Maguire, *University of Alberta*
John McLure, *University of Iowa*
Gail McLure, *ACT*
William McLure, *Iowa City*
Jennifer McCreadie, *Indianapolis, Indiana*
Les McLean, *University of Toronto*
Walter McMahon, *University of Illinois*
Colleen Medley, *Decatur, Illinois*
Betty Merchant, *University of Illinois*
Richard Merritt, *University of Illinois*

Andy Metcalf, *Illinois State Board of Education*
Barbara Metcalf, *Champaign, Illinois*
Chris Migotsky, *University of Illinois*
Cheryl Mitchell, *University of Illinois*
Carol Mills, *University of Illinois*
Louise Monegain, *Chicago, Illinois*
Linda Hamman Moore, *Parkland College*
Trudy Morritz, *University of Illinois*
Peter Mulhall, *University of Illinois*
Shankar Nair, *University of Illinois*
Susan Noffke, *University of Illinois*
Nigel Norris, *University of East Anglia*
Jeri Nowakowski, *NCREL*
Joe O`Shea, *Chicago, Illinois*
Rita O`Sullivan, *University of North Carolina, Greensboro*
Margery Osborne, *University of Illinois*
John Ory, *University of Illinois*
George Padavil, *Illinois State University*
Carmen Palmer, *Chicago, Illinois*
Shireen Pavri, *University of Illinois*
James A. Pearsol, *Ohio State University*
Mary Alyce Pearson, *Okemos, Michigan*
Michelle Perry, *University of Illinois*
Kristin Powell, *Chicago, Illinois*
Betty Braxton Preston, *EPIE*
Olafur Proppe, *University of Iceland*
Petrun Proppe, *Iceland*
John Prussing, *University of Illinois*
Laurel Prussing, *Urbana, Illinois*
James Raths, *University of Delaware*
Beverly Rauchfuss, *University of Illinois*
Thomas Rauchfuss, *University of Illinois*
George Reese, *University of Illinois*
Robert Rich, *University of Illinois*
Aimee Rickman, *University of Illinois*
Shelley Roberts, *University of Illinois*
Fred Rodgers, *University of Illinois*
Todd Rogers, *University of Alberta*
Barak Rosenshine, *University of Illinois*
Lou Rubin, *University of Illinois*
Deborah Rugg, *Urbana, Illinois*
Steve Rugg, *University of Illinois*
Katherine Ryan, *University of Illinois*
Elba Saavedra, *University of New Mexico*
Maria Saez, *University of Valladolid*

Bari Sanders, *University of Illinois*
Jim Sanders, *Western Michigan University*
Michael Scriven, *Claremont University*
Margie Skirven, *Urbana, Illinois*
Robert Skirven, *University of Illinois*
Connie Leean Seraphine, *Chicago, Illinois*
Naj Shaik, *University of Illinois*
Jonghee Shim, *Columbia University*
M. Mobin Shorish, *University of Illinois*
Ricka Shorish, *Urbana, Illinois*
Stephen Silverman, *University of Illinois*
Helen Simons, *University of Southhampton*
Kathryn Sloane, *University of Illinois*
Lou Smith, *Washington University of St. Louis*
Lu Ann Smith, *IMSA*
Nick Smith, *Syracuse University*
Ralph Smith, *University of Illinois*
Dick Smock, *Anacortes, Washington*
Theresa Souchet, *University of Illinois*
Haluk Soydan, *Uppsala University*
Bud Spodek, *University of Illinois*
Robert Sprague, *University of Illinois*
Alison Stake, *Bloomington, Indiana*
Ben Stake, *Albuqueque, New Mexico*
Bernadine Stake, *Urbana, Illinois*
Bob Stake, *University of Illinois*
Christopher Stake, *Bloomington, Indiana*
Don Stake, *Saratoga, California*
Jake Stake, *Chicago, Illinois*
Janet Stake, *Bloomington, Indiana*
Jeff Stake, *Indiana University*
Laura Stake, *Bloomington, Indiana*
Sara Stake, *Urbana, Illinois*
Bill Stallings, *Georgia State University*
Norman Stenzel, *University of Illinois*
Karen Strong, *University of Illinois*
Carolyn Sullins, *University of Illinois*
Pat Templin, *Palo Alto, California*
Paul Theobald, *Wayne State University*
Christine Thompson, *University of Illinois*
Rosalie Torres, *Alameda, California*
Penha Tres-Brevig, *University of California at Irvine*
Deborah Trumbull, *Cornell University*
Zoltan Ujhelyi, *University of Illinois*
Suzan Van Beaver, *University of Illinois*

Daniel Walsh, *University of Illinois*
Larry Weber, *Virginia Tech*
Wayne Welch, *University of Minnesota*
Tim Wentling, *University of Illinois*
Ann West, *Champaign, Illinois*
Charles West, *University of Illinois*
Ian Westbury, *University of Illinois*
Liesel Wildhagen, *University of Illinois*
Frederick Wirt, *University of Illinois*
Klaus Witz, *University of Illinois*
Eboni Zamani, *University of Illinois*
Bill Zoellick, *Boulder, Colorado*
Philip Zodhiates, *University of Illinois*

Lee Cronbach

Ernest House &
Kenneth Howe

Michael Scriven

Saturday's Opening Session
Introductory Remarks

Lee J. Cronbach
Stanford University

There is a theme behind today's session. There is a stream of ideas that go back to John Dewey. To refresh my memory for introducing Ernie House, I went back to read his paper, "Evaluation as Argument." I found a sentence on the first page which read about like this, "In a democracy, you have to assume that the people are capable of reasoning to a sound conclusion if they are adequately informed." That came straight from John Dewey.

Ralph Tyler was an admirer and associate of Dewey back in Progressive Education days and greatly influenced by him. Tyler set the pattern of evaluation for a long time. Both Tom Hastings and I were trained by him. Tom really worked in evaluation from 1942 or so through 1961 while I was off in other fields. Tom put the ideas into practice, really got the theme into the system--perhaps you would say, got it into the ideology.

I got back into evaluation by accident, when the post-Sputnick projects came to campus. Several people involved started talking to me, and I started talking to Tom not, as always, about campus gossip and ideas in general but specifically about evaluation. Tom did a great job of laying out this ideology for me.

I happened to leave Illinois exactly when Bob Stake came, but not for that reason. Nor did it deter him. Just another footnote: Years earlier, Tom had recruited me. He persuaded this University and me that we belonged together. So Tom recruited Stake. I am pretty sure that Tom would not have selected Bob if Bob hadn't already shown the democratic leanings that he has subsequently made into a central theme of his work. But I don't believe Bob had actually said much about that.

Ernie, I believe, was still in graduate school in 1963, but he joined the CIRCE team, and the three of them, Hastings, Stake, and House worked together, I don't know how closely, and I can't figure out who influenced whom, but they developed an utterly harmonious extension of these views of how to create an interface between the evaluator and stakeholders, not with the project administrators, not with the sponsors, but with the people the project served. That is precisely the theme that has brought you here today.

Now Mike. . . I really didn't tell you anything about Ernie, but I don't know that much about him. Ernie did his graduate work in administration here but he turned out unlike any other education administration thinker. Nor is he much like thinkers in the other policy lines I know of, but he certainly has been writing about policy a long time, with great originality, and that is why we continue to look up and read his early work.

Mike Scriven started out as a philosopher of science, had a glowing reputation in that field from his publications in the early fifties and early sixties that are still being cited in the philosophy literature. He got caught up because Indiana University got itself a curriculum project and decided it needed some evaluation. And, it being the fashion of the sixties that you didn't turn anything about curriculum over to the Education Department, thinking that philosophers ought to know how to evaluate, they recruited Michael to be a leader in their evaluation work. Mike did not shy off but appeared to have a moment of timidity. He said, "People have been working in this field a while. Maybe I ought to see what ideas are out there." He happened to know that I was in Education and he knew me from the work I did with Paul Meehl when Scriven was still at Minnesota, so he wrote me, saying, "What can you tell me about current thinking in evaluation?" I sent him a reprint of my 1962-63 paper, the formative evaluation piece--the "formative" term came later from Mike. That was all that Mike needed. He was so outraged by my ideas that he went on to write his famous monograph of 1965 in which he exposed these heresies of mine and made his pitch for summative evaluation.

Incidentally, this really is a reunion. Where did that monograph get published? In a series of monographs that Bob Stake organized because the AERA Executive Committee, when I was president, thought there should be such. I persuaded Bob to find a number of editors for the series. He himself edited the first volume and included "The Methodology of Evaluation," the 1965 monograph which made Mike famous in educational research. Mike held to that same theme for a long time and, as you all know, it became a widely respected view in evaluation circles.

With that, I turn the floor over to Ernie.

The Issue of Advocacy in Evaluations

Ernest R. House and Kenneth R. Howe
University of Colorado, Boulder

Eleanor Chelimsky (1998) has provided us with a valuable synthesis of what she has learned over the past decades as director of one of the most visible and highly regarded evaluation offices in Washington, the Program Evaluation and Methodology Division (PEMD) of the General Accounting Office (GAO). Of course, she is speaking from experience in one particular set of circumstances. In fact, one of her conclusions is that specific political conditions have strong effects on how evaluations are done, which suggests we should generalize to other situations with caution. Experiences elsewhere might be different.

In her article, she contrasts experience with theory, emphasizing that experience is not always consonant with evaluation theory and that theory is of dubious value. But perhaps the problem here is with what she thinks theory can provide. We develop this point in terms of what she says about advocacy, a major theme in her paper. Let's begin with experience rather than theory.

> The need in a political environment is not for still another voice to be raised in advocacy, but rather for information to be offered for public use that's sound, honest, and without bias toward any cause. Policy makers in the Congress expect evaluators to play precisely such a role and provide precisely this kind of information. . . . Yet we've seen recently attempts to rationalize *advocacy by evaluators*, and this idea has some roots in theory. . . . Our experience in PEMD was that advocacy of any kind destroys the evaluators credibility and has no place in evaluation (Chelimsky, 1998).

At the same time, she says, Congress rarely asks serious policy questions about Defense Department programs. And this has been especially true with questions about chemical warfare. In 1981 when she initiated studies on chemical warfare programs she found that there were two literatures. One was classified, favorable to chemical weapons, and presented by the Pentagon in a one-sided way to Congress. The other was critical, dovish, public, and not even considered by Congressional policy makers.

On discovering this situation, PEMD conducted a synthesis of *all* the literature, she says, "which had an electrifying effect on members of Congress who were confronting certain facts for the first time." This initial document led to more evaluations, publicity, and eventually contributed to the international chemical weapons agreements.

This chemical warfare work was predicated on analyzing the patterns of partisanship of the previous research, understanding the political underpinnings of the program and the evaluation, and trying "to integrate conflicting values" into the evaluation--which she recommends for all such studies. This is a very intelligent approach, it seems to us. Our question is, what framework guided her to conduct the study in this fashion? No stakeholder group was inciting her to do so. The Pentagon pushed its own information, and the anti-chemical doves theirs. Chelimsky had to have some framework, intuitive though it might be, for guiding her as to what to do.

We don't know what she use but we think the framework could be something like this: Include conflicting values and stakeholder groups in the study. Make sure all major views are sufficiently included and represented. Bring conflicting views together so there can be deliberation and dialogue about them among the relevant parties. Not only make sure there is sufficient room for dialogue to resolve conflicting claims, help the policy makers and media resolve these claims by sorting through the good and bad information. Bring the interests of beneficiaries to the table if they are neglected. How the PEMD evaluators accomplished all this we are not told.

Now all of this analysis and interpretation requires many judgments and decisions on the part of the evaluators as to who is relevant, what is important, what is good information, what is bad, how to handle the deliberations among policy makers, how to handle the media, what the political implications are, and so on. The evaluators unavoidably become heavily implicated in the findings, even if they themselves don't formulate the actual conclusions of the study. Their intellectual fingerprints are all over the place.

There are several points to be made here. First, she has a definite framework from which she approaches the problem, even if this framework is implicit and intuitive. Otherwise, how was she guided in what she did? Second, this framework was a combination of facts and values melded together. How others valued chemical warfare had a lot to do with how she interpreted and handled their claims. Similarly, Stake (1995) in his study of an elementary school in Chicago combines facts and values. He begins his case study by describing the school, the principal, what the teachers are doing, etc. By the time he finishes his description of Harper Elementary school, the reader knows what Stake thinks about Chicago school reform. Is this description? Yes. Is it evaluation? Yes. It is both melded together. Furthermore, the claims are objective in the sense Stake can be right or wrong about the school and Chicago reform.

To return to Chelimsky's evaluation of chemical warfare, her entire evaluation is guided by her particular conception of the role of evaluation in public policy. Is this advocacy on the part of the evaluators? We would say no, even though their work is heavily value-laden and incorporates judgment. It is not advocacy, such as taking the Pentagon or the dove's side of the issue at the beginning of the study, and championing only one side or the other. After all, if the Congress is so heavily slanted towards the Pentagon, it would make canny political sense to keep on their good side since they are the clients. Presumably, this is what client oriented evaluators (e.g., Patton, 1996) would have done. Or, they might have constructed value summaries endorsed by Shadish et al (1995), "If you are in favor of chemical weapons, X is the action to take, but if you are opposed, Y is the action to take," and turned these over to policy makers.

But the evaluators did something more risky and more defensible--they included all sides, not just the Pentagon side, in the study. This was the proper thing to do, in our view. Now it seems to us that the conduct of this study is consistent with theory, not opposed to it. Or at least the theory we want to endorse. We suggest three criteria for evaluations to be properly balanced in terms of values, stakeholders, and politics, in what we call the deliberative democratic approach

(House and Howe, forthcoming). First, the study should be inclusive so as to represent all relevant views, interests, values, and stakeholders. No important ones should be omitted. In the chemical warfare case, the views critical of chemical warfare programs were omitted originally and only the favorable Pentagon views were included, thus biasing conclusions in the previous studies.

Second, there should be sufficient dialogue with the relevant groups so that the views are properly and authentically represented. Getting authentic views is not always easy to do for various reasons but it is often critical. "Paying attention to what the beneficiaries of a program think about it is a hallmark of a credible study, and has nothing to do with advocating for those beneficiaries" (Chelimsky, 1998). In this case the potential victims of chemical warfare can hardly be present. Someone must represent their interests. Presumably including stakeholders and talking to them when possible is *not* advocacy in Chelimsky's view.

Third, there should be sufficient deliberation to arrive at proper findings. In this case the deliberation was long and productive, involving evaluators, policy makers, and the media eventually. We are not told details. Deliberation might involve ways to protect evaluators or others from powerful stakeholder pressures, which can seriously inhibit discussion, as Chelimsky notes. Proper deliberation cannot be simply a free-for-all among stakeholders. If it is, the powerful win; deliberation is aborted.

Designing and managing all this involves considerable judgment on the part of the evaluators. And we see no way around it. One can be guided by intuition, as Chelimsky and her colleagues seemed to be, or try something more explicit, as we are suggesting in our deliberative democratic approach. Actually, Chelimsky does advance a conception of the public interest, i.e., that the evaluation should be judged by "its success as a provider of objective information in the public interest."

And she goes further: "My guess is that the much greater risk to our field is not lack of use for the right reasons, but rather a declining capability or willingness to question the status quo, which is our most important task and the best justification for our work." Here she is correct in pointing to much current theory which does indeed support the status quo, however implicitly. Such theories incorporate what we call the "received view" of values, an incorrect view, as it turns out (House and Howe, forthcoming).

So isn't she an advocate for her particular conception of the public interest and of evaluation's role in it? If not, how does this view differ from advocacy? Advocacy in one sense means taking the views or interests of one group and always championing them over others, regardless of the findings of the evaluation. For example, Chelimsky and her colleagues could have taken either the views of the Pentagon or those of the doves without balancing out the two. This would be one kind of advocacy. She hasn't done this.

On the other hand, if advocacy means using or endorsing *any* particular frameworks or values, she might be accused of advocacy for her particular conception of the public interest, one not everyone would agree with. In fact, she says all evaluators should conduct evaluations with informing the public interest in mind. She might be an advocate in that sense of endorsing an overall framework. We believe that all evaluators must embrace *some* conception of the public interest, of democracy, and of social justice, even if these conceptions are implicit. They cannot avoid it in the conduct of their studies.

In *this* sense evaluators should be advocates--for democracy and the public interest--and for what this presupposes--an egalitarian conception of justice. In our view the public interest is not static and often is not initially identifiable, but emerges (or ought to) through properly constrained democratic processes in which evaluation plays a role. Interestingly, because evaluators *should be advocates* for democracy and the public interest, they *should not be advocates* for particular stakeholder groups in which perceived interests are viewed as impervious to evidence and are promoted come what may. (Greene, 1997, uses the sense of advocacy one way and Chelimsky, 1998, the other, unfortunately talking at cross purposes.) Nor should evaluators play the role of neutral facilitators among advocates of competing "value summaries," or stakeholder "constructions," in our view.

How does this chemical warfare case differ from evaluation of social programs? Not much, except in the particular views and stakeholders involved. In Madison and Martinez's (1994) evaluation of health care services on the Texas Gulf Coast, they identified the major stakeholders as the recipients of the services (elderly African-Americans), and the providers of the services (mostly white physicians and nurses), plus representatives from African-American advocacy groups. Each group had different views, with the elderly saying the services were not sufficiently accessible, and the medical providers saying the elderly lacked knowledge about the services.

Is it advocacy for particular groups, let us say the African-Americans, to include them in the study? We think it is not advocacy, but rather balancing out the values and interests of the study. All perspectives should be represented--the democratic view--and evaluators should try to determine who is correct. Nor is it advocacy to enter the study with the understanding that African-American views are often excluded in such studies. That is documented history, and the evaluator should be alert to such contingencies.

In such an evaluation, there is no grand determination of the rights of elderly African-Americans versus those of white professionals in society at large. That is beyond the scope of most evaluations. Evaluators must determine what is happening with these services in this place at this time, a more modest task. Advocacy in the misdirected sense would mean that one enters the study already convinced that the African-Americans are right and the service providers wrong, or vice versa, regardless of the facts. This is not the proper role for evaluators.

Our notion of the public interest in evaluation is one of deliberative democracy in which the evaluation informs public opinion objectively by including views and interests, promoting dialogue, and fostering deliberation directed towards reaching valid conclusions. Objectivity is supplied by inclusion, dialogue, and deliberation and by the evaluation expertise the professional evaluator brings to bear. Evaluators cannot escape being committed to some notion of democracy and the public interest. The question is how explicit and defensible it is.

References

Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation*.

Greene, J. C. (1997). Evaluation as advocacy. *Evaluation Practice*, 18 (1), 25-35.

House, E. R. & Howe, K. R. (forthcoming). *Values in evaluation and social research*. Thousand Oaks, CA: Sage Publications.

Madison, A. & Martinez, V. (1994, November 3). *Client participation in health planning and evaluation: An empowerment education strategy*. Paper presented at the American Evaluation Association, Boston.

Patton, M. Q. (1996). *Utilization-focused evaluation* (New century ed.). Thousand Oaks, CA: Sage.

Shadish, W., Cook, T., & Leviton, L. (1995). *Foundations of program evaluation*. Newbury Park, CA: Sage.

The Meaning of Bias

Michael Scriven
Claremont University

Introduction. It is a pleasure to be reunited with many old friends on this occasion. Lee's reference to the four of us makes me think that the most valuable part of it has always been the willingness of Lee and Ernie and Bob--and, I hope, myself--to challenge accepted doctrine and to be open to those who challenge the part of it that we favor most strongly. Of course, the second is the hard part of it, the love of criticism.

It is a hard for most people to realize fully that the love of criticism is an essential part of professionality. Twice in my life I have been called by prospective clients saying, "We think we've got a pretty good program here. We're not dead certain of that, even though it has had some good evaluations. We'd like you to come and shoot it up." On both of those occasions, I did so, and on the second occasion, the invitation turned out to be a lie. This was an evaluation of a computer-based approach used by the counseling center at the University of California at Irvine. It was a straightforward enough evaluation, once you took seriously the idea that the program was supposed to be providing a service to students. Doing that, I ran three of my graduate students through the program, and its disastrous failings emerged readily.

From the administrator's desk, dazzled by the computers, these failings--of content as well as of the machinery--were invisible. In any case, they refused payment in order to not have my critical report in their files. I said I would be happy not to charge them and instead use it as the theme for my next published article. So they called and said they had appointed a negotiator. I called the negotiator and asked if he was empowered to negotiate to the full amount of the contract and he said, "Absolutely." So I said fine, that I would not charge them since they did not think it worth paying for, but I would use the example in every future speech that I made on a related topic. It is a common sin to try to deceive evaluators, but an especially unattractive variant involves lying to them about your interest in criticism, faking what is perhaps the most valuable of all values in a professional.

Love of criticism is indeed a rare thing to find. I could soften this position and say that even if a professional can't manage the love of criticism, they can and must manage placing a high value on criticism. (The argument for this is via the premise that professionals must commit to lifelong learning, and that, for obvious reasons, there is no way to identify where that learning is most needed without skilled, systematic, external evaluation.) There would still be very few professional training programs in the country whose graduates were seriously taught that value and retain any trace of it. I think that in evaluation we have some of the best role models for doing it, and I feel myself very fortunate to have had the chance for long interactions with them.

Bias 101. Today I'm going to talk about bias and make a little tribute to Bob about a topic that we've been discussing lately from somewhat different points of view. Having written about this before, I put together a short paper and deciding yesterday afternoon that I didn't like it, I've been up most of the night writing it again. So I will forgive you for going to sleep if you'll forgive me for going to sleep.

The oft-given definition of bias in the statistics and methods texts is the thermometer that regularly reads too hot. A scientist from whose domain that comes would never use that as an example of bias. He or she would simply say the instrument is inaccurate or reads high. Bias is not any systematic error. Its core meaning in common parlance is a culpable human disposition to systematic cognitive error. If one wants to use it of inanimate objects, the use is by analogy, and the paradigm example of bias is the bowling ball used in lawn bowling. It is weighted--the term commonly used is in fact "biased"--so that it will roll in a curved path, deviating from the straight path that would be there without the bias.

What is actually called the bias, in the (lawn) bowling ball is in fact the lead weight in the ball that gives it the disposition to roll in a curved path. This case might be called the purely descriptive sense of bias. It's just a fact that the ball is biased. It's not an evaluative term because the error is only metaphorical, the factual deviation from a straight path (when launched in the conventional way). But it is clear that the property of bias in this case is a dispositional property.

We can establish that it is present when the bowl is made, long before it ever manifests the bias. The bias is not the deviation from the original path, but the propensity to so deviate; the disposition to deviate. This distinction between bias as systematic error, by contrast with the disposition toward systematic error, is not a mere terminological point. It is a vital point which makes possible remedial procedures in evaluation, which otherwise would be completely impossible, as you'll see. The presence of bias can be taken into account in practice by skilled (lawn) bowlers so that we are able to place the biased ball on the green exactly where we wish. Indeed, we can make it do tricks that an unbiased ball cannot do, such as hooking in behind a blocking ball. Bias in this purely descriptive sense is what makes bowling interesting. But bias in the evaluative context is itself an evaluative term, referring to the disposition to avoidable error and its presence is then by definition undesirable. It's not a desired part of the game.

The distinction between bias and the systematic error it tends to produce is critical in evaluation because it creates the possibility of controlling bias without having to remove it. And, with most biases, it's easier to control than remove. If bias were the actual bad result, it would often be impossible to remove.

While in evaluation we try to eliminate bias, we often have to settle for controlling its effects. It is frequently remarked, with some truth, that we are all biased about some things. Unfortunately, it is often erroneously concluded from this, partly because of the failure to distinguish between the bias and the systematic error it tends to cause, that there is no point at all in pushing for objectivity, since we're all biased. But objectivity in expressed views and reports is a matter of avoiding manifest bias, the effects of bias, and that there is a great possibility of controlling.

The reason for valuing objectivity, otherwise known as the absence of prejudice, is simple. Objectivity involves fewer errors. Bias, the lack of objectivity, is by definition a predisposition to error, and thereon rides the distribution of health, welfare, and happiness. It would be hard to think of a more significant reason, a better reason, for wishing to improve our qualifications in the objectivity dimension.

I was reminded the other day, of this failure to distinguish between the disposition and the results of the disposition, when the diversity officer of my university came up with a recommendation for the whole faculty to undergo diversity training. The diversity training version to which he referred involved breaking into small groups in which we would reveal our biases to each other. And then, having revealed those biases in the semi-confidentiality of a group of people, people you either don't know at all or not well, we would have cleared our minds of such wickedness, at least partially. Having cleared our minds, we would then be able to reassemble and address such matters as how to enlist more blacks as faculty members.

Now that's a typical mistake. It's a plausible mistake if you think the correct remediation model is: "Let's attack the error, get rid of it, and then everything after that will be fine." However, it would be hard to find a more naïve conception of the operation of the human mind. The problem is that, in the first place, we're not likely to reveal a racial bias, either because we're nervous about doing so or because we're not very good at identifying it in ourselves. On the other hand, given the present PC climate, we are also unlikely to take the attacks we'd get for saying that we're not biased; and who can prove differently? Great choice between unattractive alternatives!

And, in the second place, even if we did confess bias, it's not at all clear that we can voluntarily get rid of it or even significantly affect it. Certainly not by mentioning it in an arbitrary group of acquaintances. So this seems to me to be thoroughly confused, trashy pop psychology. It appears likely to be personally distressing to every honest person present, without offering the slightest chance of improvement. And it leaves the real problem still ahead of us.

What we need to be doing, certainly on my campus and I think more generally, is to be looking very hard at the recruiting procedures that we're using, how well designed they are to help with diversity, how much energy we're putting into them, and then exactly what selection/promotion procedures we're using and how well justified those are. That is, we should operate first and foremost at the action end of the problem not at the propensity end. This is where evaluation training can do a great deal to improve the situation, which is extremely badly thought out at the moment.

For example, we need to push very hard to get ethnicity and gender treated as criteria of merit in the many cases where there is a need for diverse role models in the department and a need for input from colleagues with a diverse point of view. It is a logical fallacy to describe this as reverse discrimination; that phenomenon, which certainly exists, is simply the mirror image version of the standard type of discrimination. What is being supported here is justified selection, no more, no less. We need to back away from quotas and from legal locks to named minorities, and move to a needs-based system; and understand that, properly used, the ethically defensible part of what is often called affirmative action survives in an intelligent needs-based, race-blind, gender-blind approach. The society has needs, students and potential students have needs, the campus and its components have needs, and these needs make it absurd to practice the traditional types of discrimination; and just as absurd to practice the reverse kind, which has been creeping up on us.

It's not that we should abandon the basic academic quest to appoint on merit; it's that merit isn't as simple as being good at research or good at teaching, or good at team work, or good at student counseling. It's a combination, part of which is having the talents that are needed now and in the future within the event horizon of each appointment. Speaking fluent Spanish, for example, is in many situations a job-related skill even in a mathematics department in a California college today; being female or black is just the same in many departments. It doesn't override subject-matter ability, but it sure does count as job-related, on any valid personnel evaluation approach.

While it's laudable to continue to try to reduce personal biases, and appropriate to feel bad about continuing to have them, it's far more important and much more realistic to eliminate biased selection, biased promotion, biased allocation, and biased dismissal. Most of us have the good will to make changes, but we lack the capacity to make changes that bring ourselves to a total lack or significantly improved state of bias. We need to bring our good will and brains to bear on our practices. Even if we do this, we may go to our graves secretly suspecting that we're still somewhat racist and sexist. The society can live with that as long as we keep it to ourselves and successfully control its translation into action. It is very difficult to get rid of it, and so efforts at elimination by confession are not the effective way to go. They don't work, or don't work to any degree that has been documented, whereas controlling the manifestation works very well, although of course not always perfectly.

Alcoholics are never cured, according to AA doctrine, but they can and ought to stop drinking. That's good enough. Racist evaluators will be around for a while, black and white, perhaps for most of the million years it took to get xenophobia into the genes as a survival characteristic. But racist practices, in employment and in the presuppositions of evaluation reports, are a relatively rare event these days, and should be made an exceptionally rare event. They're not gone, they're not forgotten, but they are severely restrained. As Stafford Hood reminded us yesterday, there are potential, indeed probable, elements of racial bias in our practices not thoroughly explored and dealt with yet. We still have a job to do in the elimination of bias itself. But this is not job #1. Job #1 is getting the bias out of action and practice.

In the terminology I would like to use, we should work hard on mapping the components of bias. What we must strive to eliminate absolutely is the effective component. We should also work hard on the affective component of bias, but what we must strive to eliminate absolutely is the effective component of bias. Affective bias we try to work with, but there's no guarantee we can change it substantially. Effective bias we can and should eliminate or bring it very close to the zero level.

In particular, we must absolutely reject the suggestion that, because we have not eliminated all our affective bias, it is therefore pointless to eliminate effective bias, manifest bias. We may never eliminate racism in the head; we can virtually eliminate it in practice. Diabetics almost never eliminate the love of ice cream, but those that you know have eliminated it from their regular diet, from their eating practice. The other ones you no longer know. The survivors may not have destroyed the affect but they have controlled the practice. (But they still keep working on the affect and a few of them conquer it completely.)

Once we see that bias is only "a ghost in the machine, but a devil in practice," we can begin to look more carefully at the machinery of control. We need to do that because sloppy thinking about the concept has severely handicapped our efforts to fight bias in practice, to identify the mechanisms of control that we should be using as standard procedure. In the fight over affirmative action, for example in California, we see an issue where compensatory justice is inappropriately treated as an issue of affective bias.

The first matter, that of compensatory justice, is a matter of leveling the playing field. The second matter, the matter of eliminating the practice or effects of bias, is a matter of having referees whose practice is unbiased. Both of these are reasonable things, but they're quite different. If you level the playing field and have racist referees, you're not in good shape. If you do not level the playing field and have fair-minded referees, you're also not in good shape. Still, one needs to separate the two out carefully because the fixes are different. Neither replaces the other. Both are feasible--without reverse racism.

The machinery of bias control. The basic rule for bias control is simple: reduce the role of judgment to the minimum by the use of explicit criteria, weights and synthesis rules. This is Rule 1 in bias control. It is at the point of judgment that bias begins to manifest itself. By reducing the amount of judgment that is involved, one can reduce the amount of biased judgment; not always possible, but always to be tried. Where we have archival data, the optimal move, again always to be tried (Rule 2) is to use a regression line prediction rather than human judges, as the "clinical versus statistical" studies indicate.

On the other hand, in its place, when this is very carefully defined (e.g., face recognition and some other complex pattern-recognition tasks), human judgment can beat any computer we are in range of creating. So, judgment will often remain a necessity or the best alternative, as Bob is fond of reminding us. But we can often do a substantial amount of tidying up, of definitions, weights, and synthesis rules, and when we do this, bias will have much less power to corrupt the results.

The third principle in bias control, Rule 3, is to calibrate the judges. First, by training them on cases we know, where we know the outcomes, and then, Rule 4, by selecting the best judges from the results. When we get down to cases, we can develop some further rules.

So let's turn now to three cases, each of which makes further distinction between bias and something else.

Case 1. The difference between commitment and bias is a matter which Ernie has taken up in his discussion of advocacy. One instance is the prosecuting attorney in a rape trial in New York City assisting a woman who has been raped. This attorney has made a specialty of prosecuting rapists. She is committed to that cause. Is she biased in her view of rapists? It is not at all clear that she is; one has no grounds for claiming that she is. She might be; but not from the evidence mentioned. Suppose instead that she is the mother of three small children and is a specialist in prosecuting child abuses. Does this show that she is biased against abusers? Surely one cannot conclude this without further evidence. So Rule 5 is that commitment is not a sign of bias.

In recent medical history, an interesting case is the young West Australian doctor who was totally committed to a particular theory, the theory that ulcers are caused by a virus. He argued strongly for it. Was he biased? Not unless he was so committed as to reject counter-evidence to his theory without due care in examining the new data completely. Remember, bias is the disposition to error; for someone who is well-informed, Rule 6 says: No errors, no bias. That scientist was correct in his claim that ulcers were due to viruses. He had impeccable evidence for this view. Nevertheless, he was treated with complete disdain by his seniors in the West Australia medical establishment. He was a paradigm example of an objective researcher while they were biased in reviewing his theory. Both sides were committed, only one side was biased. Commitment is not bias.

One caveat: there are special situations where one has to make a bet about where commitment will lead. Credibility is important in evaluation, not just validity. Commitment can cover a bias and, Rule 7, when we don't have a track record to show differently, and we have to take the fail safe route in order to protect innocent parties, we may sometimes be best advised to exclude those with commitments, especially public commitments. For example, in choosing a judge for a hearing on a controversial issue, one has to make a decision: should we treat prior commitment as grounds for exclusion in an area where it is important not just that justice be done but that it be seen to be done? We often play it safe and exclude judges because of family connections.

We extend this to jurors. We speak of their conflict of interest or possible bias. These considerations apply to evaluators. It is important to exclude oneself from doing summative evaluations, at least when substantial personal connections exist with anyone associated with the evaluand. But remember Rule 8: if there is not a good supply of equally competent replacement judges/evaluators, commitment is not enough to exclude relevant expertise, since it does not show bias. It is merely a weak statistical indicator of it.

Following this distinction into the evaluation field, the argument would be that it is important to avoid summative evaluation designs that are collaborative or highly interactive since it is likely that significant personal relationships will develop, such as friendship or hostility. Even if they don't, the likelihood corrupts credibility, which is often important in summative evaluation. This is not to say that collaborative or interactive evaluation designs have no place or have a less important place in the grand scheme of worthwhile evaluation-related activities. It's just that their ideal place is not in typical summative evaluations, which many of us find ourselves doing much of the time in a way that might surprise Lee.

We can do best by avoiding collaborative designs; but it does not follow that we should avoid summative evaluations when we have some views about the program's chance of success. For example, most evaluators with some subject-matter expertise in the drug abuse reduction field have some views about what kinds of approachs work and what kinds do not work, but many of them can still do a first-rate job of evaluating such a program. In this case, moving to someone ignorant of the field may cost us more in validity than we gain in credibility; the context will determine this.

Since we cannot argue that commitment entails disqualifying bias in such cases, we can only look at the track record to see if effective bias results from the affective commitment. The question is whether the evaluator is severely prejudiced, which is to say exhibits not only affective bias, but will probably exhibit effective bias as well. Is s/he unwilling to give new evidence its due weight? It is a severe condemnation of a person to suggest they are prejudiced to that degree. And it in no way follows from the fact that they think previous research indicates that one approach is more favorable than another, that they would be immune to evidence that points in another direction.

So, Rule 9, one should make such views known in advance to provide an opportunity for protest and a discussion of the situation. This is the procedure that the National Research Council follows, and seems about as good as we can get. In some areas, it is clear that almost anyone who is moderately well-informed about the area is going to have some views about the direction in which the research points, who the leading researchers are, and so on. In such cases, there is another bias control measure that should be used.

This is where Rule 10 comes in, which requires the use of the balanced panel rather than the virgin panel. If there is a need for experts, we should protect ourselves against the possibility of bias that goes with commitment by balancing this potentiality on the panel. We only rule out those who have demonstrated or conceded their inability to treat new arguments or evidence on their own merits. Note that this is not correctly described as balancing bias, but balancing the potential for bias. How do we identify judges who are severely prejudiced? From past experiences with them or from running a calibration exercise, as previously recommended. In these, we set scenarios and simulations that are closely matched to the case in which we are interested.

Case 2. (Each of these cases will get shorter and shorter, you will be pleased to hear.) Preference is not bias. There are many preferences that make it almost certain that one will select in a certain way. And this way may, in one sense or another, mistaken or erroneous. For example, the person who predictably chooses to settle back in the couchpotato attitude rather than going out and jogging around the block is undoubtedly making a mistake that is, in some sense related, to his or her health. That is predictable error, but not generally considered to be a bias. It is a preference. So we want to be sure we have more than mere evidence of statistical trends. Tendencies and choices do not illustrate bias, except after a long-chain inference. In areas where tastes rule and involve no unethical consequences, preferences are not biases. But, just as certainly, there are great areas of human interaction where bias is not merely a disposition to error, but a disposition to moral error.

We have previously distinguished between merely empirical sense of bias, as in the bowling ball, and the evaluative sense of that term. Now we need to distinguish between two evaluative senses: The epistemological sense and the ethical sense of bias. With each, we increase the likelihood of error and more than error in a factual sense. In the ethical sense, we increase the likelihood of ethically improper behavior. The paradigms of racism, sexism, and religious prejudice, all fall in this category. Moral error occurs when panelists having conflicts of interest serve on an expert panel, such as a panel reviewing applications for research funding. Some of our earlier examples fall under this heading too. Some improper behavior by evaluators falls under the category of unethical behavior. Not just being factually incorrect, but because of bias, leading to factually incorrect results.

Case 3. Last case, and this one should wake you up some. Invalidity is not bias. The last distinction I want to make will, I hope, drive a final nail in the coffin of the idea that systematic error is bias. We are all aware that some tests are biased. But there are also tests that exhibit systematic error without being biased. In order to establish this, let me suggest one rule for identifying invalidity in tests. There are several others. This rule is that a test is invalid if the standard method for scoring, if the rubric, awards points in a way that does not correspond to the merit of the performance. If the rubric awards points randomly for example, we would say that the test using that rubric is invalid. We might say it has large random error. But we would also say the same thing if the test has systematic error. For example, if the test rubric involves systematic error and awards half the points for an irrelevant skill, such as the use of calligraphy in a math test.

Given that principle of invalidity, then, all multiple choice content tests are invalid. They award 25% of the final score for blind guessing, an irrelevant skill. This is an emperor's new clothes kind of point. We've all gotten so used to multiple choice tests that despite their well-known limitations it seems absurd to call them all invalid. However, invalid they are, as normally scored. And the error in this scoring is systematic as well as very large. It is also quite easily corrected by changing the rubric to introduce negative points for serious errors and allowing partial points for near misses. That combination will produce an expectancy score of zero for blind guessing. That is the correct score for a blind guesser. Of course if you don't like this point, ignore it and cue on the earlier ones: the difference between manifest error and bias, the difference between preference and bias, the difference between commitment and bias. Those should be enough to lay to rest the textbook definition of bias as systematic error.

Commentary on Ernie House and Michael Scriven's Presentations

Lee Cronbach
Stanford University

Basically I agree with both positions that we have heard this morning. I think they have been soundly argued. It will serve us best if I speak quickly and give a slightly different view. I have to pick rather narrowly from within their presentations to find something to challenge. Well, not quite challenge, but for which to offer a different context.

In Ernie's paper, it is the statement that facts and values cannot be regarded as separate. I think we all could write essays defending that. I am going to say there is a different way of looking at the proposition. And in Michael's presentation, it is the idea that we ought to be reducing judgments to a minimum. And it is not because I am a defender of judgments but because I think they are indispensible for the questions that cross the evaluator's desk.

If we come in at the very beginning of the evaluation, as far as I know, all of us would urge evaluators to go to all the relevant stakeholders, experts, anyone, to identify questions worth asking in the field, including what to look at and what probes to use. In other words, before designing the evaluation, get candidate questions from the widest range of informants possible. I don't think anyone in this room would disagree, except as to what is reasonable or practicable.

At that point, judgments become very important. You have to decide which of these suggestions to take seriously. You have to prune the list. You have to allocate resources. You have to make some of them the focus of work, and, for some, accept much less accurate answers, and ignore the rest. And it will include judgment of the politics, such as how much difference it is going to make if you can get this matter clear, and your judgments of probability that some of the implied contentions are valid.

Now you can inform your judgment by talking to some more informants, but sooner or later, the person who signs off on the design has to approve a flock of judgments. In order to do it, he is going to have to use values, and therefore the facts you collect are heavily influenced by the persons who make these decisions, even if they do not personally state the evaluation questions or collect the data. So at that point, I agree with Ernie.

I have been thinking about the questions that are on the table today from a rather different angle. Sam Messick has caused a considerable stir in the testing community, arguing that we should consider not only the validity of the interpretations from a scientific point of view of what the test is measuring, or the implications of that, in the factual realm, *but also* the consequences of using the tests, that is, the validity of the policy of putting the test into practice.

It now seems to me that these two have to be thought about rather differently. And yet I am not satisfied with my thinking. The literature in the testing field has generally treated validity as something we testers and scientists ought to thresh out to the point that we are as sure as we can be about a) what sentences are true and b) what uncertainty should be attached to a lot of sentences that we are going to continue to use until we get better information or better theory.

That is the task for an expert community. And Mike's proposal for a balanced panel of experts makes sense to me in the evaluation context. My friends at Stanford and I wrestled with the question of getting the proper deliberations going. It is something that Ernie is pressing for. The best model that we could come up with would be something like a Royal Commission or a National Academy panel that would go through the material from an evaluation and say what are reasonable interpretations of it. It seems to me that that is sometimes viable but it is in no way manageable over the whole range of evaluations that we do.

I feel dissatisfied with that answer but I don't think anyone has offered a satisfactory one. At least it handles Chelimsky's question of the evaluator as advocate. Ernie is challenging Chelimsky. She is proposing that advocacy of conclusions is problematic. Ernie is advocating attention to certain issues, but not the conclusions she was talking about.

Ernie is an advocate of a process of analysis and education of the people using the evaluation results. Fine.

The process that Ernie is advocating requires that the evaluator be successful at managing a quick review of a mass of material and produce a definitive report. It is very different from a normal process of science that works at its own pace through an invisible college. That cannot be accelerated.

But now we turn to Messick, who is going to have us do something about the consequences of a policy of, in his case, adopting the testing. This of course is central in most illustrations of evaluation, including, for instance, the affirmative action that Michael talked about. The forthcoming edition of the Test Standards, assuming it is not changed from here on radically, handles the point that I am now coming to by just saying flatly that the Standards are going to stop with the scientific interpretation of the testing and not deal with consequences. Consequences are important but not part of the validity of using a selection test routinely, mechanically, without judgment. As for the consequences it has for eliminating certain populations from the group served, important, but not part of test validity. I can sympathize with that. It gets them out of the trenches, but not out of the remaining problems of what to do about consequences. At this point, I don't think a balanced panel is the answer.

The judgment of well qualified persons is not democracy. It is the people who have to make the mistakes. This is back to Dewey. If informing the people is not to the point where you can grab their attention, lay out before them all the alternatives, then the answer is not advocacy of what the evaluator likes. Then the public will be swayed by power. It can be best still to act through democracy. As I see it now, the choice among consequences, for example, how long, in certain situations, do you want prison sentences to be; how far do you want the University of California to lead in the direction of getting enough doctors to serve urban Black neighborhoods. These are things that the people have to come to in their own good time. It is not for the evaluators to decide where they ought to come. It is the people's decision.

But acknowledge the evaluator's job, getting the facts on the table, getting the strategic values out in the open, so that the problems are confronted. And of course, this is precisely what Bob Stake has been working on for a very long time. That is very different from trying to eliminate judgments. It says that in the long run, judgment is the function of democracy.

Linda Mabry

Jennifer Greene

Nick Smith

Deborah Trumbull

Assessing, Evaluating, Knowing

Linda Mabry
Indiana University

Introducing a panel composed of James Raths,
David Hamilton, Sue Noffke, and Gene Glass.

Good morning. I'm Linda Mabry. I was a doctoral student at CIRCE from 1987-1991, at that time not much interested in the topic of our next session, "Assessing, Evaluating, Knowing" --or, at least, not much interested in assessment and evaluation.

But these are matters close to the heart, close to the bone of the Stake social science agenda, and they became matters of enduring fascination for me personally.

Let's begin our session, shall we, with a couple of innocent questions --

"Is that true?"
"Who says?"

Turns out, we don't know what is true, if anything is true.

"What can we **know?"**

In a constructivist age, in which everyone is understood to be constructing individual understandings, truth is idiosyncratic.

My truth is not your truth. I have warrants for my truth, evidence and reasons which persuade me. But so do you.

"Who says?"

Turns out, it doesn't matter. Even Catholics believe the pope is fallible.

In ages past, we granted religion a monopoly on Truth. But there are different religions, and they tell different truths.

Christianity tells of Adam and Eve, but the Native Americans of the Northwest say a raven pecked open a clamshell and discovered the first human beings. (I rather like that.)

Science tells it all differently--big bangs and a series of odd mutations.

In the modern age, we granted science a monopoly on Truth. And there have been other ages. But we are here--scientists all, social scientists.

What do we know, what *can* we know when we evaluate a program?

Our descriptions of the programs we evaluate *mis*represent, turn people into "stakeholders" -- personnel, beneficiaries, "impactees[1]," or (worse) stats.

We tell *our* side. Even when we try to tell *their* side, it's from *our* perspective.

The many values dear to the many stakeholders are so diverse we cannot select or devise standards by which to evaluate program quality without neglecting or offending some.

And who are *we* to decide which representation is "accurate"? ... which standards to apply? ... which interests to prioritize? ... to whom to tell our "truth"?

How dare we claim such authority or reinforce the idea that there *is* a truth, a *reality* about a program, a feasible, reasonable, proper *way* to evaluate it?

Evaluators all know from experience that clients often feel bitterly *mis*represented by a negative evaluation, that Scriven's standards and procedures can yield a dramatically different judgment of program quality than Eisner's or Stake's.

Are Eisner and Scriven wrong, and Stake right? Or the reverse?

What do we know, what *can* we know when we **assess** student achievement?

---

[1] Term used by Scriven earlier in the day.

Even those--most of us in this room--who have great test scores somewhere on our dossiers, are not likely to think that a score from one uncomfortable morning says much about who we are, what we know, what we can do.

Our hero of the day, for instance--Did you know Stake passed French as a graduate student at Princeton? and that he'll be making his final comments today in French?

(C'est vrai, n'est-çe pas, mon ami? Tu parle français?) Or that he deliberately flunked a math test?

(Was that for the Navy?)
(What was it you were trying to get out of?)
(Am I leaving out any good parts?)

Our miscreant has been lucky. But not everyone caught up in assessment is lucky.

And it's not just standardized, norm-referenced, multiple-choice tests that are a problem.

In February, a teacher in a rural middle school in Pennsylvania told me, as she was preparing for her students to take a state-mandated *performance* assessment in writing:

"This test is not really a fair representation of how well my students can write. They could do better if they had a choice of topics and could do the sorts of things we usually do [in class] . . .

"I'm frustrated. We've done all this preparation and I've organized to the max so they can concentrate on their writing during the test time, but my students will still score at only about the state average.

"We're a rural district, and we don't have all the curricular options and resources you find some places . . . But my students will be compared to students who are in suburban schools with a writing-only curriculum.

"When the scores are printed in the newspaper, people will think we're not doing a good job of teaching here.

"It just doesn't seem fair."

(Pennsylvania middle school teacher,
personal communication, February 11, 1998)

Last month, a principal in northern Michigan talked to me about that state's High School Proficiency Test which is *not* a requirement for graduation. He said:

"Last year, our best students--kids with straight As--were declared 'non-proficient' on the basis of this test.

"So, this year, about a third of our parents exempted their kids from the test so they wouldn't have a black mark on their transcripts. In one classroom, not a single student showed up.

"It's the brightest kids who are getting exempted, so I'm expecting a big disaster."

"When they print the scores in the newspaper, people will only look at the numbers--'Numbers don't lie!' If I try to explain, they'll think I'm making excuses.

"I just hope the neighboring districts have more exemptions than we do."

(Michigan high school principal,
personal communication, April 29,
1998)

Fortunately, we have good people here (today) to help us think about these matters. Jim [Raths], have I said anything true? What do you say?

Balancing Philosophy and Practicality in Qualitative Evaluation

Jennifer C. Greene
Cornell University

What I have tried to do for this occasion is to make some connections between the teachings of Bob Stake and those things that currently trouble me.

A Bit of History

The recent history of evaluation, especially social and educational program evaluation in the US, is well known. The significant contributions of Bob Stake's theories, thoughts, and lifework to the course and the temperament of evaluation over the past 30 years are also well known. In brief:

> 1. Bob Stake, along with other visionaries of that era (notably for me, Lee Cronbach), helped fledging evaluators such as myself in the mid-1970s, first, to make sense out of the *mismatch* between what I knew how to do--experimental designs and statistical analyses--and what was likely to be meaningful to those in the sites in which I was working (in those days, many schools and other educational sites with ESEA Title I, Title III, and Title IV grants).
>
> 2. Second, Bob Stake and others also helped fledging evaluators such as myself to begin to learn about *alternative* ways to do evaluation and other forms of applied social inquiry, alternatives that relied on (a) a different worldview, a grounded, constructivist set of philosophical assumptions about our social world and how we can know it; (b) a whole new set of methods intended to capture the meaningfulness of people's experiences in qualitative, not numeric, form; and (c) the idea that evaluation could and should be *responsive* to people in the settings in which we worked, in addition to remote decision makers. In this way Bob Stake significantly contributed to the *direction and the course* of contemporary program evaluation.

> 3. *And* third, Bob Stake and others helped us to start to develop our self-consciousness about how our methods make statements about *values*, about the value choices available to evaluators, and about the challenges of honoring multiple value stances and perspectives in any given evaluation. In this way, Bob Stake significantly contributed to the *temperament* of contemporary program evaluation.

With these influences, I and many other fledging evaluators of that time re-educated ourselves, reframed our work, and set off on our new course, that being the course of qualitative, constructivist evaluation, and that being a course that was guided by a value-conscious temperament and specific values like responsiveness, usefulness, integrity, and fairness.

Now, some 15-20 years later, we have survived the paradigm wars, we have refined our own theories and thoughts, methods and manners, and we have claimed a secure place for qualitative approaches to evaluation. The challenges continue, however, challenges both to the essential nature of qualitative evaluation and to its role and voice in social policy and program decision making. Let's hear some of these challenges.

Challenges From The Center

Selected statements from the center evaluation community--on what we as evaluators should be doing these days . . . As the center, these ideas still constitute the dominant discourse and therefore are difficult to simply ignore.

> *Joseph Wholey*:
>
> At a time of severely constrained resources and declining public trust, the Government Performance and Results Act and related performance initiatives offer exciting opportunities for evaluators to help improve government performance and help restore public confidence in government. . . . Current reform efforts will increase the demand . . . for evaluability assessment, outcome monitoring, interrupted time series studies, and qualitative evaluations of the effectiveness of public programs and of the reform efforts themselves. The demands will present

> exciting political, bureaucratic, and technical challenges for evaluators (1997, pp. 129-130).
>
> *Robert Granger* (MDRC):
>
> Evaluators [must] attend to the need for sufficiently credible counterfactuals at all stages of their work. Doing so . . . require[s] that they develop strong theories, use multiple methods of inquiry to search for and confirm patterns in data, creatively blend research designs . . . [and] inevitably . . . confront the test of trustworthiness . . . [for which] random assignment has been characterized as "the gold standard" or "nectar of the gods" (1997, pp. 5, 19).
>
> *Eleanor Chelimsky*:
>
> [Today there is an] overriding need for evaluation credibility . . . mean[ing] a judgment by others . . . that the evaluation is both competent and objective. There are, in fact, a great many things we can do to foster both objectivity and its appearance, not just technically, in the steps we take to make and explain our evaluative decisions, but also intellectually, in the effort we put forth to look at all sides and stakeholders of an evaluation. . . . *What seems least well understood . . . is the dramatically negative and long-term impact on credibility of the appearance of advocacy in an evaluation* (1997, pp. 58-59, emphasis added).
>
> *Michael Scriven*:
>
> It is my contention . . . that both distancing [staying at arm's length from those being evaluated] and objectivity remain correct and frequently achievable ideals for the external evaluator, ideals to which we must try to adhere as closely as possible even when circumstances put full realization beyond our grasp. . . . Tempering validity with mercy . . . is a violation of validity--and validity is the highest professional imperative of the evaluator, as of the radiologist or engineer or historian (1997, p. 483).

In other words . . .

Joseph Wholey continues to promote evaluation as technical service to government. Differences among us all can fit within our large and ever-expanding toolkit, which offers tools for all occasions. Today, says Joe, our toolkit can be especially useful in helping government agencies to meet GPRA requirements.

Bob Granger exemplifies the evaluative understanding of many who conduct national-level evaluations of significant social interventions--in the domains of education, job training, housing, community development, and now welfare reform. And, even more importantly, this experimentalism captures the evaluative thinking of many of our public decision makers.

Eleanor and Michael steadfastly and loyally continue to honor truth and its disciples of objectivity and neutrality as guiding ideals for evaluation.

But, to Joe Wholey, ask those once-fledging evaluators enlightened by the teachings of Bob Stake and others, isn't evaluation much more than a set of techniques and evaluators more than technicians? And where in your toolkit is there room for philosophical differences and especially value consciousness?

But, ask we to Bob Granger and to Eleanor and Michael, isn't the richness of human experience inadequately understood, even diminished, by the experiment? Isn't the very meaning of truth contested by different philosophical stances and contextualized by the vast diversity of lived experience? And, how are the interests of all stakeholders, especially those on site and those usually not heard, really served by obeisance to objectivity?

There is much at the center of evaluation that remains at odds with the interpretive, contextual, responsive *direction* and value-conscious *temperament* that Bob Stake has contributed to our field.

But, I feel these *pulls from the center*, sometimes strongly, and they have led me to want more from my qualitative convictions. As I wrote recently:

> Qualitative evaluators have importantly helped to educate decision makers about the idiosyncratic, deep and inherent complexities of human phenomena. . . . But, offering too much complexity can immobilize those charged with making decisions, and, at times, qualitative evaluators have done just that. [Further] qualitative evaluators . . . [have] reject[ed] objectivity in favor of celebrating subjective insights and knowledge claims, and ... discount[ed] the relevance of existing theory and past research in favor of a "grounded and emic" understanding of a particular context, [and so] qualitative inquirers have

> become good storytellers. Good stories illuminate the human condition, but don't usually offer specific solutions or recommend alternative endings, each based on different value stances and perspectives. [Perhaps] it is time for qualitative evaluators to do more than tell good stories; it is time for them to reclaim their full responsibilities as scientific citizens (Greene, 1998, p. 141).

It is these kinds of *pulls from the center* that have realigned my antennae towards actively seeking out other challenges and alternatives. I wish to speak with more authority and I am looking for help in doing so. So, as counterpoint to the center, I ventured out to the edge and looked at the contemporary discourses of postmodernism, feminism, critical social science, and other edge inhabitants. These discourses include but are not exclusive to evaluation. And they offer many challenges of importance to qualitative evaluators.

Challenges From The Edge

I have sampled here three of the many challenges from the edge: challenges to the very nature of our qualitative data, challenges to the meaning of our interpreted meanings, and challenges to the political location of our work.

Challenges from the edge regarding the very nature of our qualitative data . . .

> *Jim Scheurich*:
>
> The [qualitative] interview interaction is fundamentally indeterminate. The complex play of conscious and unconscious thoughts, feelings, fears, power, desires, and needs on the part of both the interviewer and interviewee cannot be captured and categorized. In an interview, there is no stable "reality" or "meaning" that can be represented. The indeterminate totality of the interview always exceeds and transgresses our attempts to capture and categorize. When we think we "interpret" what the meaning or meanings of an interview are, through various data reduction techniques, we are overlaying indeterminacy with the determinacies of *our* meaning-making, replacing ambiguities with [*our*] findings or constructions. When we proceed as if we have "found" or "constructed" the best, or the key, or the most important

> interpretation, we are misportraying what has occurred. . . . [Instead in the analysis] the researcher fills [the interview's] indeterminate openness with her or his interpretive baggage; imposes names, categories, constructions, conceptual schemes, theories upon the unknowable; and believes that the indeterminate is now located, constructed, known. Order has been created. The restless, appropriative spirit of the researcher is (temporarily) at peace (1995, p. 249, emphasis added).

Camille Tischler:

Like Jim Scheurich, Camille Tischler highlights the indeterminacy of qualitative data and then pointedly critiques our coding and categorization analytic techniques because they:

- "fail to address the complexities of human discourse" (1997, p. 2)
- fragment and decontextualize the holistic unity of experience (p. 3)
- treat interview data as "primarily information transfer," rather than as an intentional and relational human exchange (p. 2)
- and thereby, fail to acknowledge the gap between language and meaning (p. 5), and
- especially fail to include the relational dimensions of human interaction and experience

In other words . . .

These authors contend that neither our qualitative data themselves nor our analytic ability to find meaning in these data can be warranted. Rather, all qualitative inquirers (or any other social inquirer, for that matter) can do is reveal the indeterminacy of human interactions and experiences. Jim then advises us to acknowledge our own baggage of biases, our positionality, to the fullest extent possible, and also to "foreground the open indeterminacy of the interview interaction itself" (p. 250) in our work and our reports of our work. Camille advises us to re-emphasize the narrative, the story, as a better (although still flawed) standard for our work.

Challenges from the edge regarding the meaning of our interpreted meanings . . .

> *Leslie Goodyear*:
>
> Postmodernism opens space for new forms of representing social science inquiry by challenging the assumptions of what are seen as accepted forms of presenting the findings of inquiry. [Marianne] Paget (1995) points out, "there is something odd about privileging an analysis of discourse in its least robust form, a written text, exploring it in great detail while ignoring the speakers' miens and intentions . . . (p. 229)." By allowing for many possible interpretations of events and texts, postmodernism also creates an intellectual space where [from Patti Lather] "data are used differently; rather than to support the analysis, they are used demonstrably, performatively."
>
> In the creation of new representations of inquiry, we need to struggle to represent the complexities and indeterminacies of participants' experiences . . . [and] to acknowledge our role in the construction of the representation, our voice in the presentation. [Further] as, in postmodern terms, knowledge is partial, conditional and contextual, so are representations (1997, pp. 64-65, p. 69).

In other words . . .

Our reports, as representations of indeterminate meanings, are themselves indeterminate and therefore should be "interrogated" or questioned as to form, authorship, and meaning, both as presented and as received.

Challenges from the edge regarding the political location of our work . . .

> *Michelle Fine and Lois Weis*:
>
> How do we handle "hot" information, especially in times when poor and working-class women and men are being demonized by the Right and by Congress? . . . For instance, what do we do with information about the ways in which women on welfare virtually have to become welfare cheats to survive? *("Sure he comes once a month and gives me some money. I may have to take a beating, but the kids need the money.")* A few [of those we study] use more drugs than we wish to know . . . some underattend to their children well beyond neglect. . . . To ignore these data

> is to deny the effects [of hard economic times]. To report the data is to risk their likely misinterpretation. In a moment in history when there are few audiences willing to reflect on the complex social roots of community and domestic violence and the impossibility of sole reliance on welfare, or even to appreciate the complexity, love, hope, and pain that fills the poor and working class, how do we display the voyeuristic dirty laundry that litters our database? At the same time, how can we risk romanticizing or denying the devastating impact of the current assault on poor and working-class families launched by the State, the economy, neighbors, and sometimes kin (1996, pp. 258-259)?

In other words . . .

Michelle and Lois confront head-on the dilemmas of the "public intellectual," particularly at the "hyphen" between scholarship and activism. They agonize over the risks of reporting versus not reporting the data they have, of withholding data that confirm society's worst stereotypes about the character of poor people versus distorting society's full understanding of what life is like for a poor person today. In good postmodern form, they wish to dissolve these and many other dichotomies and instead "float across" once-rigid boundaries towards new places and spaces of being.

Some Reflections

My journeys to the edge, as exemplified by this sampling, did not yield ideas and insights about how to claim greater authority, voice, and scientific citizenship in my work as a qualitative evaluator.

Instead, my journeys yielded magnificent challenges to my voice and to any authority I might once have thought that I had. Say these challenges from the edge--not only can I not claim greater voice in public decision making, but even my contextualized and partial voice as storyteller is but a fleeting glimpse of human indeterminacy, conditioned by the form of the story (or the play or the video or any other representation) that I choose to tell *and* by whoever is listening. And I must be sure to be careful about who *is* listening, because some will surely distort and co-opt some parts of the story as I wish to tell it, in which case I may want to tell it differently, or perhaps not.

So . . . I experience *pulls from the center* to claim a stronger voice for qualitative knowing and understanding, for more authoritative stories about the complexities of lived experience--stories that can carry more power than R^2's and F statistics on the average difference in something measureable between experimental and control groups.

And, I experience *pulls from the edge* that deconstruct the very concepts of voice and authority, that weaken and condition any of our claims to know anything, that primarily offer more questions and doubts than answers (and even that, in their *extreme or skeptical* form, from Linda Mabry, 1997, offer fatalism, nihilism, and ultimately, only disengagement).

HELP!!!!

More Reflections

A plea for help in resolving or escaping from this dilemma is probably hopelessly modern, as postmodernism rejects dualisms in any form. Yet, not requesting help is hopelessly or skeptically postmodern; it's giving up, it's disengaging. As educational researcher Mark Constas recently said:

> Postmodernist culture has produced . . . in the same breath an invigorating and paralyzing skepticism" (Terry Eagleton). . . . Perhaps a state of temporary paralysis was needed to make the educational research community pause and examine the assumptions and political consequences of its work. Still, *the paralysis is a state from which we must recover* . . . . We must not forget that education is about the possibility of growth and the realization of human potential. . . . We must, therefore, continue to question the value of emergent paradigms, especially those that displace pragmatic ideals so central to education (1998, p. 32, emphasis added).

And, fortunately, there is some help in recovering from this paralysis, much of it from fellow evaluators. Here is a brief sampling.

Practicing evaluation postmodernly. One, we can abandon our modernist struggles to resolve this dilemma--of conflicting pulls on our evaluation theory and practice--and learn to live with its ambiguities and uncertainties; in Tom Schwandt's words (1997a, p. 102), "accepting incredulity and doubt as modal postmodern responses to all attempts to explain ourselves to ourselves." We can become postmodern in our evaluative work.

Tineke Abma (1997a, 1997b, in press) has offered us wonderful, even inspirational examples of postmodernly evaluation that is affirmative and positive (Mabry, 1997). For example, Tineke promotes in her work the idea and experience of *playfulness*. She says, "a playful person is not too attached to his or her personal persuasions and appreciates the power of redescribing, the power of language to make new and different things possible and important" (1997a, p. 44). Tineke also invokes the "self-reflexive, polyvocal, and multi-interpretable" (in press, p. 2) texts of postmodern writers in endeavoring to craft her evaluation reports as "open, ambiguous, and unpredictable . . . without summary, conclusions and recommendations" (1997b, p. 106) and thereby as invitations to dialogue (in press).

(See also Stronach, 1997, and Stake, 1997 for thoughts on postmodernly evaluation practice.)

Seeking still other emergent paradigms, philosophies, frameworks. Two, we can search out still other paradigms, philosophies, and frameworks to guide our work.

Back to the center, there is the work of Pawson and Tilly (1997), and more recently, Henry, Mark, and Julnes (in press), on *emergent realism* as an alternative paradigm for social inquiry. Emergent realism focuses on the multiple layers and levels of explanatory mechanisms of human behavior and thereby offers room for both macro and micro perspectives, both generalizable and situational understandings, and other once-incompatible dualisms.

From the edge (or perhaps the side?), Tom Schwandt (1989, 1997a, 1997b, 1998) has for over a decade offered *practical philosophy* as an alternative way of conceptualizing the practice and discourse of social science, including evaluation.

> Practical philosophy is concerned with the mode of activity called the practical (praxis). Its subject matter is how an individual conducts her or his life and affairs as a member of society (1998, p. 9).
>
> It yields practical knowledge, which is "action-oriented self-understanding" (p. 10) rather than knowledge of how to make something.
>
> Praxis is embedded within a tradition of communally shared understandings, values, commitments, and principles vitally connected to one's life experience (p. 14).

To practice evaluation within this new frame of practical philosophy means to radically shift from a methodological to a political-ethical frame, to resist the assimilation of evaluation praxis (deliberation, practical activity) to technique (method), to be less concerned about perfecting the validity of our methods and more concerned about helping practitioners to deliberate well, to develop their own wise practice.

Promoting activism and advocacy. Three, we can concentrate on the political, ideological dimensions of our various philosophies, and even more importantly, on the political, ideological dimensions of our evaluation contexts, and actually use our work to do something about it. We can become advocates and activitists in our work.

(The proponents of critical race theory are doing just that. The recent issue of Qualitative Studies in Education (volume 11, number 1, 1998) is itself an education in critical race theory.)

Critical ethnographers Michelle Fine and Lois Weiss argue that "researchers can no longer afford to collect information on communities without that information benefiting those communities in their struggles for equity, participation, and representation" (p. 271). They continue:

> We try to position ourselves self-consciously and hope that our colleagues who are engaged in critical work . . . will enter with us into this conversation about writing the rights and wrongs in the field. . . . Many of our colleagues, on both the Right and Left, have retreated to arrogant theory or silly romance about heroic life on the ground. Others

> meticulously and persuasively deconstruct the very categories we find ourselves holding on to in order to write a simple sentence about community life. We toil on, looking for friends, writing for outrage, searching for a free space in which social research has a shot at producing both social theory and social change as the world turns rapidly to the Right (pp. 271-272).

Toiling on

I like these words because I agree with them. I believe we *are* activists and advocates in our work and need to more clearly and assertively claim these roles. And I believe that "toiling on" in the sense of claiming our own voice through action importantly honors the legacy of Bob Stake. Evaluation that toils on is evaluation that engages the meaningfulness of human travails and glories, that revels in the moment while seeking to transcend it, and that is anchored in an appreciation for what connects us together despite our vast differences. Evaluation that toils on is evaluation that strives to be philosophically thoughtful and coherent, *yet ultimately privileges* the gritty human struggles and needs, the essential human experiences and interactions, the urgent human demands and requirements of the practical context. This is the sense of toiling on that Bob Stake has taught us. Thanks Bob.

References

Abma, T. A. (1997a). Playing with/in plurality. *Evaluation, 3*, 25-48.

Abma, T. A. (1997b). Sharing power, facing ambiguity. In L. Mabry (Ed.), *Evaluation and the postmodern dilemma. Advances in Program Evaluation: Vol. 3* (pp. 105-119). Greenwich, CT: JAI Press.

Abma, T. A. (1998). Writing for dialogue. *Evaluation*, 4(4), 434-454.

Chelimsky, E. (1997). The political environment of evaluation and what it means for the development of the field. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 53-68). Thousand Oaks, CA: Sage.

Constas, M. (1998). The changing nature of educational research and a critique of postmodernism. *Educational Researcher, 27* (2), 26-33.

Fine, M., and Weis, L. (1996). Writing the "wrongs" of fieldwork: Confronting our own research/writing dilemmas in urban ethnographies. *Qualitative Inquiry*, 2, 251-274.

Goodyear, L. K. (1997). *A circle that it's time to open: Using performance as a representation of a participatory evaluation.* Master's thesis, Policy Analysis and Management, Cornell University, Ithaca, NY.

Granger, R. C. (1997). *Establishing causality in evaluations of comprehensive community initiatives.* Paper prepared for the Aspen Roundtable on Comprehensive Community Initiatives for Children and Families, Aspen, CO.

Greene, J. C. (1998). Qualitative, interpretive evaluation. In A. J. Reynolds & H. J. Walberg (Eds.), *Educational research for educational productivity. Advances in Educational Productivity* (Vol. 7, pp. 135-154). Greenwich, CT: JAI Press.

Henry, G. T., Mark, M. M., & Julnes, G. (in press). Emergent realism and evaluation. *New Directions for Evaluation.* San Francisco: Jossey-Bass.

Mabry, L. (1997). A postmodern test on postmodernism? In L. Mabry (Ed.), *Evaluation and the postmodern dilemma. Advances in Program Evaluation* (Vol. 3, pp. 1-19). Greenwich, CT: JAI Press.

Pawson, R. and Tilly, N. (1997). An introduction to scientific realist evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 405-418). Thousand Oaks, CA: Sage.

Scheurich, J. J. (1995). A postmodern review of research interviewing. *Qualitative Studies in Education, 8*, 239-252.

Schwandt, T. A. (1989). Recapturing moral discourse in evaluation. *Educational Researcher*, 18 (8), 11-16, 34.

Schwandt, T. A. (1997a). Whose interests are being served? Program evaluation as a conceptual practice of power. In L. Mabry (Ed.), *Evaluation and the postmodern dilemma. Advances in program evaluation: Vol. 3* (pp. 89-104). Greenwich, CT: JAI Press.

Schwandt, T. A. (1997b). Evaluation as practical hermeneutics. *Evaluation, 3*, 69-83.

Schwandt, T. A. (1998). *Recapturing moral discourse in evaluation--revisited.* Keynote address at the annual Kelly Conference, Ottawa, Ontario.

Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 2lst century: A handbook* (pp. 477-500). Thousand Oaks, CA: Sage.

Stake, R. E. (1997). The fleeting discernment of quality. In L. Mabry (Ed.), *Evaluation and the postmodern dilemma. Advances in Program Evaluation, Vol. 3* (pp. 41-59). Greenwich, CT: JAI Press.

Stronach, I. (1997). Evaluation with the lights out: Deconstructing illuminative evaluation and new paradigm research. In L. Mabry (Ed.), *Evaluation and the postmodern* *dilemma. Advances in Program Evaluation, Vol. 3* (pp. 21-39). Greenwich, CT: JAI Press.

Tischler, C. (1997). *"A-exam" in evaluation.* Department of Policy Analysis and Management, Cornell University, Ithaca, NY.

Wholey, J. S. (1997). Trends in performance measurement: Challenges for evaluators. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 124-133). Thousand Oaks, CA: Sage.

Naturalistic Generalizations as the Source of Investigative Insight[1]

Nick L. Smith[2]
Syracuse University

Prologue

As a "student of Bob Stake's" (it's a title more than a description), I was often accosted by other students demanding to know what Bob "really means" by what seemed to them to be an obscure term, or subtle argument, or arcane example. Since Bob was my advisor, they presumed I had special access to private interpretations of Bob's thought. But, although Bob may at times be enigmatic, he is not duplicitous, and I was usually as confused as my classmates. Almost 25 years later, and I am still trying to interpret Bob's work, at this point to my own students who frequently ask, "but what does Stake really mean by that?" Perhaps Bob's greater contribution has not been the answers he has given us, as much as the questions he has challenged us to consider.

All of you who teach have probably had an overly eager student who takes one of your perfectly good ideas and enthusiastically contorts it into something no longer resembling your original meaning. Numerous times, Bob would peer over his glasses at me with his puzzled look, baffled by the meanings I could construct from his sensible words. In what follows I am once again trying to understand what Bob really means. It seems fitting, at this celebratory event, that I give Bob one more chance to set me straight--which he will undoubtedly do if he happens to be in the room.

---

[1] Revised July 15, 1998

[2] Address correspondence to Nick L. Smith, 330 Huntington Hall, School of Education, Syracuse University, Syracuse, New York, 12344-2340. 315-443-5259 (O); 315-445-0250 (H); 315-443-9218 (Fax); Email: NLSMITH@SYR.EDU.

Introduction

How do evaluators, researchers, and inquirers in general, achieve their insights? Consider two possibilities: naturalistic generalization and investigative insight.

Perhaps researchers achieve insight through what Bob Stake has referred to as "naturalistic generalization." He introduced this construct in 1980 in an American Educational Research Association annual meeting paper (Stake, 1980), followed by his 1982 article with Deborah Trumbull (Stake & Trumbull, 1982), a chapter in Ernie House's 1986 book (Stake, 1986a), and in Bob's 1995 book on case study research (Stake, 1995), as well as elsewhere. In their 1982 article, Bob and Deborah (Stake & Trumbull, 1982) suggest that naturalistic generalization provides a fundamental basis for the improvement of practice.

> Almost absent from mention is the *common* way in which change or improvement *is* accomplished, the way followed intuitively by the greatest and least of thinkers ... One may change by adding to one's experience and re-examining problems and possible solutions intuitively. . . . program evaluation studies should be planned and carried out in such a way as to provide a maximum of vicarious experience to the readers who may then intuitively combine this with their previous experiences. The role of the program evaluator or educational researcher would then be to assist practitioners in reaching new understandings, new *naturalistic generalizations* [emphasis in original] (Stake & Trumbull, 1982, pp. 1-2).

Interestingly, most of Bob's writing seems to concern how to facilitate the reader's or stakeholder's naturalistic generalizations, rather than the mental processes of how the researcher acquires his or her own understandings. But, in places, we might infer that Bob believes naturalistic generalization is a source of insight for the practice of the researcher as well. After all, the vicarious experience developed for the reader is to be based on the evaluator's own personal experience with the program. Perhaps, then, naturalistic generalization is the mechanism of insight for both the reader and the researcher.

> The dominant belief is that formal generalizations, conceptual knowledge, is the essential ingredient of improved practice. Our position is that practice is guided largely by tacit knowings, by naturalistic generalizations, formed from experiencing, often implicit. (Stake & Trumbull, 1982, p. 11)

Consider another alternative. For several years, I have been interested in the nature of investigative inquiry. I have studied the accounts of work by investigative journalists (Cornwell, 1989), forensic and physical anthropologists (Maples & Browning, 1994, Rathje & Murphy, 1992), criminal and medical investigators (Thompson, 1988, Sacks, 1995), and investigative physical, biological, and social scientists. I have tried to discern the fundamental methods by which investigators do such diverse work as find lost children (Greene & Provost, 1988), determine the cause of airline crashes (psychological and political causes as well as physical causes (Emerson & Duffy, 1990)), uncover the operations of Wall Street (Stewart, 1991) and the Internal Revenue Service (Burnham, 1989), and trace the causes and spread of epidemics (Larson, 1998).

I have suggested elsewhere (Smith, 1992) that there are methods and mental processes common to all these varieties of investigative inquiry. Common aspects of investigative inquiry include:

> (1) Investigative contexts: Both local and broad social, historical contexts are relevant.
>
> (2) Investigative purposes: The goal is to uncover something hidden, through the various roles or inquiry games played by investigators (journalists, pathologists, social scientists, and so on).
>
> (3) Investigative process: The process is problem-oriented, recursively emergent, alternatively [sic: alternately] exploratory and confirmatory, and focused on the development of lines of argument.
>
> (4) Investigative means: The methods or techniques used depend on the investigative context, the game being played, the phenomena of interest, but all investigations

> require the mental powers of knowledge, observation, reasoning, and intuition.
> (Smith, 1992, p. 10)

For the next few moments I would like to contrast these two possible approaches to researcher insights. This is my first attempt to present some of these ideas; they represent a "work in progress," but I proceed, remembering that CIRCE was always a place where people did not fear to make foolish statements (as evidenced by their frequency).

Naturalistic Generalizations versus Investigative Insights

A few important contrasts between naturalistic generalizations and investigative insights are suggested in Table 1. Neither time nor space permit a detailed elaboration of these contrasts, but a brief overview will provide a helpful orientation for subsequent discussion.

Table 1.
Naturalistic Generalizations Versus Investigative Insights

Contrasts	NATURALISTIC GENERALIZATIONS	INVESTIGATIVE INSIGHTS
Approach	Case Specific Studies	Case Specific Studies
Goal	Inferences About Personal, Subjective Phenomena	Inferences About Hidden, Unknown Phenomena
Phenomena	Events, Conditions, Actions & Meanings	Events, Conditions, Actions & Motives, Causes
Researcher Orientation	Holistic, Integrative	Analytic, Reductionistic
Researcher Claims/Assertions	Descriptions, Constructions, Understandings	Descriptions, Discoveries, Explanations, Understandings
Researcher Mental Posture	The Receptive Mind	The Probing Mind
Products	Portrayals, Vicarious Experience	Multiple Lines of Argument

The two frameworks of inquiry within which naturalistic generalizations and investigative insights operate evidence both similarities and differences. Both approaches employ predominantly case study methodology to develop inferences about initially unknown phenomena. Although both are concerned with events, conditions, and actions, naturalistic generalizations focus more on possible meanings people have or attribute to personal, subjective phenomena. Investigative insights focus more on motives and causes associated with hidden but more objectivist phenomena.

Posture and methods differ more dramatically across the two frameworks. In achieving naturalistic generalizations, the researcher cultivates a receptive mind, seeking holistic, integrative understandings, and especially constructions of meaning that can be communicated through the sharing of vicarious experience. In achieving investigative insights, the researcher cultivates a probing mind, employing analytic and reductionistic methods to discover and develop principally causal explanations that are communicated through the statement of multiple lines of argument.

Naturalistic generalizations and investigative insights appear to be fundamental products of a researcher's inquiry, but each serves a different inquiry purpose and employs those methods suitable to its particular phenomena of interest. Although there are a number of ways to explore the connections between naturalistic generalizations and investigative insights, the concept of "intuition" provides an enlightening intersection.

Intuition in Naturalistic Generalizations and Investigative Insights

Intuition is the faculty (or what I have referred to as a "power of the mind") by which we access tacit knowledge, knowledge known but in unexpressed form, knowledge one has but cannot explain how acquired. Intuition is "direct perception of truth, fact, etc., independent of any reasoning process; immediate apprehension" (Random House, 1967, p. 747). Should we be apprehensive of these immediate apprehensions of knowledge? Not according to Bob. But he has been severely criticized, for example by Denis Phillips (1987), for rhetoric, behind which ". . . lurks an epistemology that is scandalously charitable, for it lacks an explicit recognition of the need to put knowledge-claims to the test" (p. 94). "If a qualitative researcher believes that he or she has achieved 'understanding,' according to Stake, then this claim must be accepted--it is as simple as that!" (p. 93).

Well, perhaps not *quite* that simple. Bob does not advocate uncritical acceptance of intuitive insights: "In our search for both accuracy and alternative explanations, we need discipline, we need protocols which do not depend on mere intuition and good intention to 'get it right.' In qualitative research, those protocols come under the name of 'triangulation'" (Stake, 1995, p.107). He goes on to identify four types of triangulation protocols: data source triangulation, investigator triangulation, theory triangulation, and methodological triangulation.

Although these forms of triangulation enable the researcher to produce more warrantable assertions, they do not necessarily appear to be the basis of the researcher's own naturalistic generalizations. In response to David Hamilton's suggestion that naturalistic generalizations are best thought of as private knowledge, Bob says, "I agree that such generalization loses its experiential privateness even when made conscious to that same person . . . Translation from experiential language to formal language diminishes and distorts some of the meaning" (Stake, 1995, p. 86). Indeed, at this point, the researcher's naturalistic generalizations appear very similar to Elliot Eisner's (1991) connoisseurial understanding. Stake's triangulation protocols are thus analogous to Eisner's methods for moving private connoisseurship to public criticism. I'm still not clear how Bob thinks the researcher's naturalistic generalizations arise--are they private, spontaneous, intuitions?

The researcher's assertions may or may not be passed on to the reader as explicated generalizations, but do contribute to the vicarious experiences from which readers are to produce their own naturalistic generalizations. Foremost in the construction of vicarious experiences for readers, however, seems to be the researcher's own personal experiences. Stake suggests that most qualitative researchers ". . . favor a personal capture of the experience so, from their own involvement, they can interpret it, recognize its contexts, puzzle the many meanings while still there, and pass along an experiential, naturalistic account for readers to participate themselves in some similar reflection" (Stake, 1995, p. 44).

While Bob does provide guidelines for how the researcher can facilitate naturalistic generalizations by the reader (e.g., Stake, 1995, p. 87), he says little about the ethical problems of researchers possibly misleading readers, whether intentionally or through their own lack of self-awareness and skepticism. Indeed, he (Stake, 1995) encourages researchers to anticipate the effect of the vicarious experiences on the reader and to attempt to create experiences as impactful as reality itself.

> The researcher should try to anticipate what vicarious experiences will do for the reader, should try to organize the manuscript so that naturalistic generalization is facilitated. By providing information easily assimilated with the reader's existing knowledge, the writer helps readers construct the meanings of the case (p. 126).
>
> . . . Naturalistic generalizations are conclusions arrived at through personal engagement in life's affairs or by vicarious experience so well constructed that the person feels as if it happened to themselves. It is not clear that generalizations arrived at in two quite different ways are kept apart in any way in the mind. One set of generalizations through two doors (p. 85).

Such a position seems cavalier, given the serious problems in society related, in part, to individuals' difficulty in discerning the differences and implications between vicarious and personal experience--from the possible effect of the media (society's most powerful creator of vicarious experience) on violence and ethical behavior, to the possible "implanted" memories of the Repressed Memory Syndrome (Loftus and Ketcham, 1994), to the apparent self-deception of the practitioners of the controversial Facilitated Communication strategy for assisting communication by persons with autism (Burgess, Kirsch, Shane, Niederauer, Graham, & Bacon, 1998). In evaluation, I (Smith, 1990) have pointed to the problematic use of naturalistic generalization and connoisseurial evaluation in meta-evaluations, such as Bob's Cities-in-Schools meta-evaluation reported in *Quieting Reform* (Stake, 1986b). Evaluators are not immune to unwittingly creating vicarious experiences that encourage readers to share the evaluator's own biases.

Although naturalistic generalization for readers may be "... the *common* way in which change or improvement *is* accomplished, the way followed intuitively by the greatest and least of thinkers." (Stake & Trumbull, 1982, p. 1-2), it is not clear whether the naturalistic generalizations of the researcher are private intuitions, the result of cultivated expertise, or some mix of the two. In investigative insight, intuition is seen as a highly developed mental ability.

> . . . the powers of *intuition* are perhaps the category of mental abilities least often acknowledged in discussions of methodology but most often highlighted in anecdotes of investigative insight. The important role of intuition and even the conditions of its occurrence in scientific investigation have long been recognized (see Beveridge, 1957) . . . (Smith, 1992, p. 9).

Whereas Bob describes a naturalistic generalization as a more or less self-validating action which the researcher facilitates for the reader, intuition in investigative insight is seen as a continual mental activity of the researcher. Further, Bob proposes naturalistic generalization as a primary method in the context of justification for readers to construct inferences valid for them. In investigative insight, intuition about the phenomena of interest plays a more critical role in the context of discovery.

Investigative inquiry proceeds in an alternately exploratory and confirmatory, recursively emergent, manner to develop and justify claims in the construction of multiple lines of argument designed to fully explain a problem posed within a particular investigative enterprise. I (Smith, 1992) have described this process as employing the simultaneous, synergistic operation of four mental abilities: intuition, plus knowledge, observation, and reasoning:

> First, an essential aspect of any investigative activity is the prior and ongoing accumulation of *knowledge*. Knowledge about the phenomena under study is, of course, a prerequisite to, the purpose for, and the end result of the investigation. But knowledge of both the local context of the phenomenon and the broader social, historical context

> of the investigation is also needed. Further, knowledge of the game or role played by each particular form of investigation is necessary for successful participation (p. 8). . . . Each form of investigation requires different types of knowledge--both public and personal knowledge and both propositional and tacit knowledge from study and experience (p. 9).

Increased knowledge facilitates intuitive insights, while, at the same time, intuition suggests what needs to be known next and how that knowledge might be acquired.

> Second, the mental powers needed to conduct investigative inquiry include the powers of *observation*. I do not mean observation in the narrow sense of data collection but rather in the more profound sense of knowledge about what to look for, the ability to recognize the meaning and significance of what is seen, the ability to perceive and interpret. Obviously, these powers of observation presuppose much prior knowledge and experience (p. 9).

Again, intuition guides observation, just as observations provide the content of which intuitions are formed.

> Third, the powers of *reasoning* are needed for any investigative inquiry, especially when the intent of that inquiry is to build a line of argument or chain of reasoning that fully explains a problem within the confines of a particular context and inquiry game. Characteristic of investigative inquiry is the simultaneous development and testing of multiple lines of argument (p. 9).

Though rationalist constructions, lines of argument are often guided by intuition, as is the selection of relevant evidence, the means of testing claims, and the sense of when to move from discovery or exploration to confirmation and back again.

Intuition is thus a critical aspect of investigative inquiry and operates in conjunction with knowledge, observation, and reasoning to produce investigative insights. In a narrow sense, Bob's naturalistic generalization by the researcher appears to refer to the direct intuitive apperception of tacit knowledge, that is, intuition; in that sense, naturalistic generalization is a primary source of investigative insight, hence the title of this paper.

Conclusion

Again, I suspect I am not fully appreciating the subtleties of Bob's arguments. He speaks of searching for happenings and promoting empathetic understandings rather than constructing causal explanations. We are probably working at different purposes with our inquiries. I am concerned with how we might improve the human condition by understanding how things in the world around us work, while Bob, ever the teacher, is more concerned with the educative process of shared meaning. As he says, "Often, the researcher's aim is not veridical representation so much as stimulation of further reflection, optimizing readers' opportunity to learn" (Stake, 1995, p. 42). Over the years, Bob has certainly optimized my opportunity to learn, for which I am ever grateful.

References

Beveridge, W. I. B. (1957). *The art of scientific investigation*. New York: Vintage.

Burgess, C. A., Kirsch, I., Shane, H., Niederauer, K. L., Graham, S. M., & Bacon, A. (1998). Facilitated communication as an ideomotor response. *Psychological Science, 9* (1), 71-74.

Burnham, D. (1989). *A law unto itself: Power, politics, and the IRS*. New York, NY: Random House.

Cornwell, J. (1989). *A thief in the night*. New York, NY: Simon & Schuster.

Eisner, E. W. (1991). Taking a second look: Educational connoisseurship revisited. In M. W. McLaughlin & D. C. Phillips (Eds.). *Evaluation and education: At quarter century*. NSSE Yearbook, Pt. II, 169-187. Chicago: University of Chicago Press.

Emerson, S. and Duffy, B. (1990). *The fall of Pan Am 103: Inside the Lockerbie investigation*. New York, NY: G.P. Putman's Sons.

Greene, M. and Provost, G. (1988). *Finder: The true story of a private investigator*. New York, NY: Crown Publishers, Inc.

Larson, E. (1998, February 23). The flu hunters. *Time*, 54-60, 63-64.

Loftus, E. & Ketcham, K. (1994). *The myth of repressed memory*. New York: St. Martin's Press.

Maples, W. R. and Browning, M. (1994). *Dead men do tell tales*. New York, NY: Doubleday.

Phillips, D. C. (1987). *Philosophy, science, and social inquiry*. Oxford: Pergamon.

Rathje, W. and Murphy, C. (1992). *Rubbish! The archaeology of garbage*. New York, NY: Harper Collins.

The Random House dictionary of the English language, the unabridged edition. (1967). New York: Random House.

Sacks, O. (1995). *An anthropologist on Mars.* New York, NY: Knopf.

Smith, N. L. (1990). Cautions on the use of investigative case studies in meta-evaluation. *Evaluation and Program Planning, 13* (4), 373-378.

Smith, N. L. (1992). Aspects of investigative inquiry in evaluation. In N. L. Smith (Ed.*), Varieties of investigative evaluation. NDPE,* vol. 56, 3-13. San Francisco, CA: Jossey-Bass.

Stake, R. E. (1980, April). *Generalizations.* Paper presented at the American Educational Research Association annual meeting, Boston.

Stake, R. E. (1986a). An evolutionary view of educational improvement. In E. R. House (Ed.*), New directions in educational evaluation.* Philadelphia, PA: Falmer.

Stake, R. E. (1986b). *Quieting reform: Social science and social action in an urban youth program.* Urbana, IL: University of Illinois Press.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R. E. & Trumbull, D. J. (1982). Naturalistic generalizations. *Journal of Philosophy and Social Science, 7* (1), 1-12.

Stewart, J. B. (1991). *Den of thieves.* New York, NY: Simon & Schuster.

Thompson, J. (1988). *Gumshoe: Reflections in a private eye.* Boston, MA: Little, Brown & Company.

Naturalistic generalizations: We think what we are

Deborah J. Trumbull
Cornell University

A Bit of History

When I submitted the abstract for this paper, the title I chose was: *Naturalistic generalizations: We are what we think.* However, after finishing the actual paper, I switched the title to its present form. I trust that by the end of the paper it will be obvious to the reader why I made the switch.

I said I would do something on naturalistic generalizations for Robert Earl Stake's retirement celebration because I wanted to return to a paper published in 1982 in the *Review Journal of Social Science.* The paper developed because one of the editors, Nelson Haggerson, solicited a piece from Bob Stake. I was a doctoral student working with Bob at the Center for Instructional Research and Curriculum Evaluation (CIRCE) at the time. Bob pulled out a 1980 talk he had given at the American Educational Research Association, and gave it to me with instructions something like, "Here, work on this and turn it into an article." Or something equally incisive yet ambiguous. Many of us lucky enough to have worked with Bob likely have received a similar request.

After a few extensions of the deadline, the piece was published with both our names on it. I negotiated the extensions. Nelson would call CIRCE and I'd get the call and would explain that we were still working, and things were getting done, and could we have just a bit more time. What that usually meant was either that I was having a panic attack and unable to do anything, or that I had done something and given it to Bob and was waiting to hear the verdict from him and having a panic attack. I'm sure some of you have had that experience of waiting for word from your mentor.

When I'd get a verdict from Bob, I'd write some more, trying to be guided by Bob's comments and my attempts at understanding what he meant. Eventually the piece got to a point where Bob let me send it to Nelson, who accepted it and published it. Every time I've seen Nelson since then, he's praised the piece and said he continues to use it and find it helpful.

I've always felt awkward at this praise. Initially I demurred, explaining that really the ideas were Bob's, that my role, which role I chose for myself, was limited to adding some embellishments. What I added related to my own experiences as a teacher who had been involved in a significant attempt to revise my own curriculum and teaching. My small additions focused on sterility of the formal generalizations that came from the standard educational research of that time. Process-product research on teaching was searching for generalizations based on the relation between narrowly prescribed variables. The variables themselves were attempts to operationalize a few key constructs. As an experienced practitioner, I viewed these generalizations as weak and pale compared to the kinds of understanding I had developed of my work. I had a strong belief that much of the published research of the time was irrelevant and incapable of contributing to changing practice. I do not think I added as much to explicating the notion of naturalistic generalization as Bob might have liked.

Writing on the Beach in San Diego--Naturalistic Generalizations of Place

It was odd to recall these memories of the genesis of the 1982 article when I began writing this paper. I wrote while at a conference in San Diego, staying at a resort run by the Princess cruise lines. In the literature we had received, the resort was described as being on an island in a bay, with wonderful landscaping. I had images of the rocky, windswept islands in the Massachusetts bay near which I lived for several years. Then, I arrived in San Diego. From my Northeastern perspective the island of the conference was a bunch of landfill in the middle of an extensive marsh, with bright sunshine sporadically filtered through a mix of sea fog and the nacrous smog from San Diego. The landscaping was half Fantasy Island, half Jurassic Park. It just was not right. And added to that, the days of the CIRCE I was remembering–in the southwest corner of the second floor of the Education Building, with my desk tucked in with other grad students in the antechamber to Bob's office--seemed far away, and grey.

Re-experiencing CIRCE from San Diego--Naturalistic Generalizations about Writing

One reason those days of CIRCE seemed grey is that I was in a fog a lot. The difficulties I had with adding to Bob's thinking for that article were not really a function of Bob's instructions to me. The problem was that I could not hear what he said. I had no naturalistic generalizations about the task because I had no experience with the task.

The writing Bob expected me to do as a developing academic was foreign to me, so foreign I did not realize it was foreign. I, of course, wrote for courses, for CIRCE projects, and generally got reasonably good evaluations. But I couldn't quite grasp the task that Bob had set me. There were, I think, two reasons for this, which relate to my recent thinking about this notion of Naturalistic Generalization.

1) I did not understand negotiating ideas in a way that could have contributed more to the article. I was not able to grasp what someone said and then augment that through discussion, dialogue, argument, explication. I was still at the point of taking others' ideas and adding them together to create my own espoused view. I was reminded of this stage at the conference, listening to earnest young scholars cite chapter and verse from various authorities, and build up elaborate positions out of other people's thinking. Where, I wondered, were their ideas? Where were their own voices? In 1982, I had not developed my academic voice, nor did I know how one could do this. I recall noting that associate professors could publish using fewer citations than assistant professors, and that full professors could publish with very few citations at all. I associated this more with status than with the development of voice.

2) I had not learned that writing is a form of thinking, which must be fluid and evolving. The test of any writing comes only with its reading. The need for the reader's reaction gives writing a feeling akin to going down stairs in the dark, feeling unsure about the presence of the next stair tread. In writing, it is the reader who furnishes the support. This sensation--of hurling oneself into space--was very hard for me. I was used to school writing and, however subconsciously, had learned to ferret out the rules for doing things the right way. At the time of working with Bob on the article, I did not help much because I was afraid to deviate from the rules that I knew must be there, even though Bob wouldn't say what they were. So, to be safe when I couldn't figure the rules out, I added little to the thinking.

Disorientation in San Diego

Even though my contribution to the 1982 paper was not as much as I wish it had been, I continue to have faith that the notion of Naturalistic Generalization is a helpful one. Naturalistic generalizations generate expectations, as do all generalizations. We develop them in one setting and apply them in other situations, whether similar or not, using them to interpret our experiences (Donmoyer, 1990). I was so disoriented in San Diego as I wrote because the ocean was just not right. The tides moved too little in the so-called bay. The smell was wrong, the ocean too calm, too pacific. And most of all, it was in the wrong place, it was west. Growing up in the Midwest I knew my landscape by the compass points, even though I have long since translated my compass points to the East coast. San Diego's ocean reversed my compass, violated my generalizations about where the ocean should be and how it should behave.

I become Bob Stake to my students

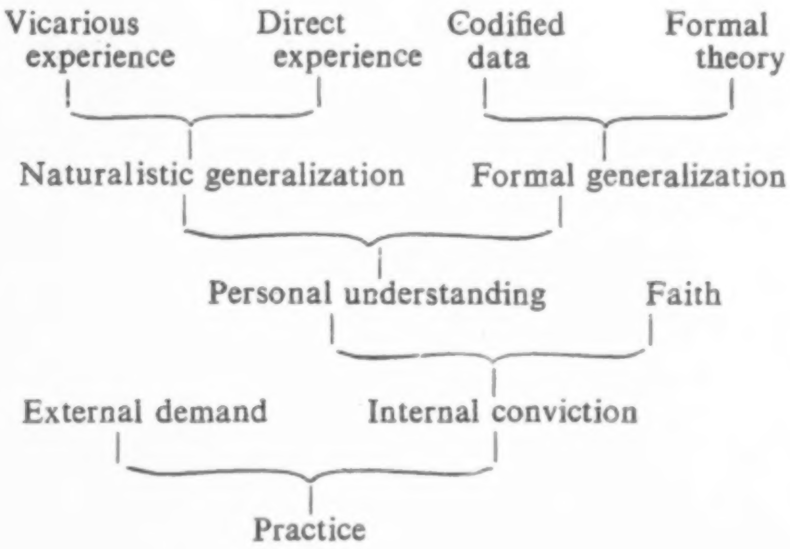
Before I went off to San Diego, I told my interpretive research class that at the stage of their projects, we (I and two grad students working with me) could not tell them what to do for their final projects. The initial class assignments had had detailed guidelines and structure, but for the final drafts, we could only tell students if they'd done well or not. The students were shocked. Silent. I got e-mails. I got assignments turned in with notes "I'm not happy with this, but there's no more time." or "I had a migraine so I'm turning in what I have right now, but I'll turn in the next version tomorrow."

Whew! I realized I was being Bob. So why was I creating a situation for my students similar to one that caused me such stress? I've learned about writing. What it is like and how it feels to present someone something I've written. I have Naturalistic Generalizations that I express in the images I use, the instructions I give, the way I manage my course. But, there are some things you just can't tell people. They have to experience them. These give rise to Naturalistic Generalizations. And so I'm trying to help my students develop these experiences, learn the writing process from the inside, develop naturalistic generalizations that will help them in their next writing assignments.

The Uses of Naturalistic Generalizations

In the 1982 article, we included a chart Bob had developed:

The Elements Of Action


Vicarious experience
Direct experience
Codified data
Formal theory
Naturalistic generalization
Formal generalization
Personal understanding
Faith
External demand
Internal conviction
Practice

As I review this chart I am struck by its appeal and how it helps to explain the **context** for the use of the term naturalistic. Someone had questioned Bob about the use of the term because it suggested the naturalistic fallacy, the belief that what is, is what **should** be. Bob, I think, chose the term naturalistic to contrast it with formal generalizations derived from experimental-style studies. He wished to explicate the bases for actual action in the world and to honor the important role played by knowledge gained from experiences and from such elusive things as faith and conviction. I will not comment on these more elusive, though intriguing, aspects of the chart.

However, I think it important not to romanticize generalizations derived from one's experience because they can engender a resistance to change: "It's always been this way, and it can't change because it has to be this way."

Return from San Diego and The Art of Case Study Research

After returning from my disorienting time in San Diego, I reread Bob's current book, *The Art of Case Study Research* to focus on references to Naturalistic Generalization. I'd known the references were there when I proposed my paper, because when I got my copy of *The Art of Case Study Research* I first thumbed through the index looking for my name and had found it in the sections on naturalistic generalization. When I continued with this paper after San Diego, I experienced a moment of angst similar to that which I'd had when working on the first paper in the early 80s. As I reread *The Art of Case Study Research* I first uttered the Homer Simpson response, "What he said." Bob's diagram was gone, somewhat surprisingly because I continue to find it compelling, but there were many amplifications of ideas I had continued to think about during the 14 years after leaving CIRCE. Bob focuses on how the writer of a case study should seek to engage the reader to the degree that the reading should be capable of generating Naturalistic Generalizations. He complexifies the notion of reader, and the active role of this reader in developing her own interpretation of the case, of the rich data presented by the researcher. He mentions the roles of the case study researcher, including advocate, teacher, biographer, evaluator.

I read with increasing chagrin. How would I move into the expected academic discourse for this paper, the "Yes, but" response? My quandary led me to wonder if I had made absolutely no progress since my student days. Why did I most immediately want to honor my sense of relation with Bob by saying, "Yes, me too, I agree with what my mentor has written, with how he's developed his argument." I thought further, though, and surfaced an aspect of naturalistic generalization that I felt Bob still had not addressed to the degree I would. The first segment of this aspect involves the relation between naturalistic generalizations and feelings. In the 1982 article is a quote from Thomas Flanagan in *The Year of the French*: "We posses ideas, but we are possessed by feelings. They lie too deep for understanding, astir with their own secret life and carrying us with them" (Stake and Trumbull, 1982, p. 7).

By making this contrast, Bob seemed to want to separate naturalistic generalizations from feelings, keeping them more in the realm of ideas. I believe, though, that naturalistic generalizations are inextricably linked with feelings. And thus, our naturalistic generalizations carry us with them. For example, under certain conditions sound travels great distances over water. I remember learning about this phenomenon in a physics class and finding the explanation intriguing because I knew this was an actual phenomenon, having experienced it myself, and the conditions under which the phenomenon occurred called to mind old feelings. I recalled going to bed as a child in the long summer evenings in my bedroom by a lake, falling asleep listening to the sound of the drive-in movie that was far on the other end of the lake. It was the evocation of those rich memories from childhood that gave richness to the spare lines in the physics explanation. Had I not had those experiences, had I not known one could expect sound to travel in that way, I would not have cared to understand the physics explanation. In this case, the naturalistic generalization, the expectation that sound would travel over water, was true. For once, the physics textbook amplified my own experiences.

Uses of Naturalistic Generalizations

I agree that naturalistic generalizations are a part of an individual's personal understanding, are not articulated, and have developed through experiences. They shape our expectations for what will happen and our explanations of what has happened. They can be surfaced and examined, though never completely. The ineffable is an ineluctable part of our knowing. [That was a sentence Bob struck from my dissertation. Ha! Finally, I get to use it. But, to the point.] As Gudmundsdottir wrote:

> Hirsch uses the metaphor of an iceberg to describe the two kinds of activities that constitute interpretation. The tip of the iceberg is explicit interpretation, which is what we say the things mean, and what we write in our research reports, clearly documented using quotations from data. The biggest part of the iceberg, however, is submerged and out of sight. That corresponds with informal or implicit

> interpretation (Gudmundsdottir, 1996, p. 301).

Our naturalistic generalizations are embedded in the submerged part of the iceberg. They shape how we interpret in ways that we will not be fully aware of. In *The Art of Case Study Research* Bob nods to the knotty epistemological and ontological issues in research by identifying three realities. The first is an external reality, the second is "a reality formed of those interpretations of simple stimulation, an experiential reality representing external reality so persuasively that we seldom realize our inability to verify it." The third is "a universe of integrated interpretations, our rational reality" (p. 100). Realities two and three are understandings reached by each individual, "but much will be held in common" (p. 101). Bob's examples through this section refer to the moon, the stars and the sky, arthritic knees and images of a grandfather walking with canes, and crossing the street in traffic. These are all images we share in common, or could easily consider ourselves sharing.

The Social World and Naturalistic Generalizations

I believe, though, that a more careful consideration of naturalistic generalizations requires a social constructivist viewpoint as we consider doing the various genres of research better. We develop naturalistic generalizations through our experiences in the world. The social world in which we act is more re-active, though, than the moon. We humans are born into physical *and* social worlds. The expectations of these two worlds limit what we can do, but in different ways. At the weights of humans, gravity on the earth affects us all similarly. The social world does not limit us all equally, or about the same things. Some humans early learn that they must not cry or be seen to cry. Some are taught they cannot admit to feelings of vulnerability. Some are taught not to speak of their accomplishments. Some are taught not to challenge their mentors. Some are taught it is *crucia*l to challenge their mentors. Some are taught not to acknowledge they have been mentored. We develop these naturalistic generalizations and believe in them, trust them. All of these understandings frequently remain as naturalistic generalizations, unexplicated and unexamined. They simply reflect the way the world is, because this is the way the world has been in our experiences.

But within the social structures that shape our worlds, there are different positions, and these positions are not equal. Those in some positions are not allowed to participate to the fullest of their talents, not allowed to be a fully contributing member of a society.

Positionality

It is imperative for us as researcher to examine our actions to look for the operation of naturalistic generalizations that have developed through our experiences of privilege, or powerlessness, of aspects of our position that enable us to understand the world in a certain way. If we, as researchers, are to contribute to a democratic society, it is key for us to examine ways our understandings are partial, ways our naturalistic generalizations have been formed from our positions within existing social structures.

My angst about writing this paper was occasioned--at least partly--by a conflict between my wish to honor Bob in the ways I was raised as a woman to honor, and the ways I have seen successful academicians honor. The differences between these ways has a lot to do with gender socialization, of course. The emotion engendered by this academic task--what could be more stereotypically academic than honoring someone by submitting ourselves to two days of papers, delivered when the Illinois spring weather was outstanding--alerted me to some naturalistic generalizations that I have so far failed to explicate completely. Acting in a way that was counter to my naturalistic generalization about the proper procedure for honoring someone about whom I care generated feelings. My feelings were not free-floating, they were tightly tied to ideas that, when unexamined, possessed. me. As researchers who hope to engender change, we must realize that to challenge our own and our readers' naturalistic generalizations, expectations about how things are, will engender emotional reaction, whether of relief or anger or compassion. As we attempt to understand how someone with a very different position in the social world interprets her world, we will challenge naturalistic generalizations about how the world is, both our own and our readers. We must invite that challenge, and the subsequent emotional reaction.

References

Donmoyer, R. (1990). Generalizabilty and the single-case study. In E. W. Eisner and A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate.* New York: Teachers College Press.

Flanagan, T. (1979). *The year of the French.* New York: Holt, Reinhart, and Winston.

Gudmundsdottir, S. (1996). The teller, the tale, and the one being told: The narrative nature of the research interview. *Curriculum Inquiry, 26*(3), 293-306.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks CA: Sage Publications.

Stake, R. E. & Trumbull, D. J. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science, 7*(1&2), 1-12.

Ulf Lundgren

Jim Raths

Saville Kushner

What Is Really At Stake?

Ulf P. Lundgren
The National Agency for Education
Sweden

As I am not a native master of the English language it is to take a risk to play word games. But what I am going to share with you invites doing it. To learn a second language is most of the time to learn to master a language outside its living context and that opens up for associations as well as misunderstandings.

The question, what is really at stake, contains two levels or two dimensions. On one hand it alludes to what I personally have found in so intriguing, but also provoking in the thinking of Robert E. Stake. On the other hand the title refers to what happens with education in general and educational evaluation in specific. Embedded in the presentation is the fact that I am translating my thoughts not only into another language but also into another context.

The Countenance of Educational Evaluation

I first met Bob and Bernadine in the early seventies. For me as a graduate student Bob was a mastermind picturing the face of evaluation, not the least in the article *The Countenance of Educational Evaluation.*[1] At that time educational evaluation was serious business in Sweden and thus the hope of a future for a young researcher. It was not only, to talk with Bob, *"President Johnson, President Conant, Mrs Hull (Sara's teacher) and Mr. Tykociner (the man next door)"* that had faith in education. In Sweden all had faith in education. And above all even if we had different ideas of what education is, we (and especially the politicians) shared the belief that education had to be evaluated. The progress of education was a question of rational decisions based on evaluations. I shared the same beliefs and Bob's structuring of the field of educational evaluation gave comfort.

---

[1] Stake, R.E.: *The Countenance of Educational Evaluation.* University of Illinois: Center for Instructional Research and Curriculum Evaluation. 1966.

Let me just briefly explain this deep belief in the good of evaluations.

Already during the war in 1940 the first committee was established with the task to reorganise the Swedish educational system. The aim was to increase the level of education in order to meet a "knowledge society" and to prepare the coming citizens for a democratic society. In doing that it was important to organise the school system in such a way that it gave equal opportunities. The basic question was formulated around ability grouping. The 1940 committee could not agree on the organisation and was followed by a parliamentary commission in 1946 that drew up the main lines for the educational policy to come. Still the problem around the organisation was not solved. The question at the time for ability grouping was given different answers. One way to find an answer was to move the question from the political arena to the arena of science.

The idea was to have an experimental period with different organisational solutions and by evaluations create a basis for a decision. Very few evaluations were carried out, only one main study was done--the Stockholm study.[2] In 1962 a comprehensive school was implemented and the National Board of Education was given the task to continuously evaluate the school system and from these evaluations suggest adjustments and changes--the continuous Curriculum reform. The first suggestion for a Curriculum change came in 1969, which trigged off a lively debate about national evaluations. The National Board of Education was criticised for not having fulfilled its task. In the Parliament of 1970 education was in focus and voices were heard for the forming of an independent institute for national evaluations. Three years earlier, in 1967, Urban Dahllöf published a reanalyses of the Stockholm study in which he showed that behind small differences in outcomes from comprehensive schools and streamed schools there were striking differences in time spent.[3] From this analysis Dahllöf formed the Frame Factor Theory, which I later developed.[4]

---

[2] Svensson, N-E.: *Ability Grouping and Scholastic Achivement.* Stockholm: Almqvist & Wiksell. 1962.

It was in this heated climate Bob and Bernadine first landed in Sweden. Bob's outline of the countenance of educational evaluation fitted well into the debate on the role of evaluation for educational progress and with the theoretical model by Dahllöf. We were on speaking terms and that is rare with masterminds.

The Vegetable Beef Stew

Later the whole Stake family came to Sweden for a sabbatical at the University of Gothenburg. Bob had changed from the solid organiser of models of evaluation to the doubter of models. His papers were hard to grasp, just to mention one title *"The Vegetable Beef Stew."* Titles that are not that easy to understand even for a native born master of English.

What Bob opened up was a new discourse on educational evaluation. It was a discourse that questioned established views and most of all established methods. The nearly eternal debate (which I still do not understand) on quantitative and qualitative methods had started and in Sweden a hermeneutic perspective found its place in social scientific research.

---

[3] Dahllöf, U.: *Skoldiffrentiering och undervisningsförlopp.* Stockholm: Almqvist & Wiksell. 1967. Dahllöf, U.: *Ability Grouping, Content validity and Curriculum Process Analysis.* New York: Teachers College Press. 1971.

[4] Lundgren, U.P.:*Frame Factors and the Teaching Process. A contribution to curriculum theory and theory on teaching.* Stockholm: Almqvist & Wiksell, 1972. In part published in: Donald E. Orlosky,D.B. & Othanel Smith, B: *Curriculum Development: Issues and Insights.* P. 31-34. Chicago: Rand McNally Publishing Company, 1978. Compare Lundgren, U.P.: *Model Analysis of Pedagogical Processes.* Lund: Liber Läromedel/CWK Gleerup, 1977. 2:nd ed. 1982 In part published in Giroux, A. N., Penna, W. F.: *Curriculum & Instruction. Alternatives in Education.* Berkeley, Calif.: McCutchan Publishing Corporation, 1981. And in: Giroux H. & Purpel, D.: *The Hidden Curriculum and Moral Education. Deception or Discovery?* Berkeley, California: McCutchan Publishing Company, 1983. See also: Frame Factors and the Teaching Process. In: The International Encyclopedia of Education. Research and Studies. Vol. 4.P. 1957-1962. Oxford: Pergamon Press, 1985.

In the middle of this heated debate on the use of evaluation, on methods for evaluation and models for evaluation Bob formed new ways of thinking and new strategies, which of course was annoying for my firm rational beliefs. But something happened and these new ideas found their place and had an impact on the discussion on national evaluation in Sweden. The concept of evaluation had been focused on outcome variables that could be quantified and compared on the same scale. In comparing the outcomes of two alternatives--Is A better than B or in other words is ability grouping better than non-ability grouping? Such questions could be answered or believed to be able to answer by sound statistical models and quasi-experiments.

But when the main organisational decisions had been taken, the questions from policy makers to evaluators and researchers were much more complicated and demanded new ways of understanding the role and methods of evaluations. The demands on the quality of education from the stakeholders become more articulated, demands that could not be met by national statistics and results from measurements only. The concept of educational evaluation had to be widened and questioned. The case study methodology formed by Bob became one answer to how national evaluations could be supplemented.

The process of education came in focus. New models were formed as responsive evaluation as well as new metaphors introducing new ways of defining quality. The Cambridge manifesto expresses clearly these currents and is one memorial in the history of educational evaluation.

The Storehouse of Models and Methods

But the development of the field of educational evaluation was not only a question of methodology and respondents to education it was also a question on what questions that in fact could be answered. Having the belief that evaluations can improve the national standards of education it was important to find answers on rather complex questions such as how to value equal opportunities. I have--I am sorry to say in vain--tried together with Sigbrit Franke[5] to argue against Bob that the question of methodology is subordinate to the question of what answers you want to construct, i.e. the theoretical aspect of educational evaluation. And here I do think we are facing the complex of translating not only language but contexts as well. Theory based evaluation can associate to scientific models like the ones from the early seventies, but we alluded to broad systems filled by imagination, history and culture. I am sorry we never met on that point, but there is still plenty of time.

In the eighties the concept of evaluation was widened and a storehouse of models and methods was built.

And in the eighties, once again, presidents, teachers of our children and our neighbours expressed faith in education. The media society flowered and neo-liberal solutions searched for problems. Education is always suitable for identifying problems. It is in many places the biggest local industry. Everyone has an experience; everyone has and ought to have an opinion. Opinions can be exploited, developed and extended.

The structure of production changed and thereby changed the labour market. New demands on education were formulated and a new "knowledge society" was claimed.

Most industrial societies went through educational reforms.[6] In the United States I can see by the development of standards as a movement towards centralisation. In Sweden with a highly centralised educational system the move was towards decentralisation and the forming of an independent school system. The possibility to choose and to exit was in focus articulating demands from parents well educated by earlier educational reforms. To change from a centralised system to a decentralised system and keeping the basic ideology of a school system providing equal opportunities national evaluations once again was the focal point. The National Board of Education was replaced by the National Agency for Education for which I was given the responsibility not only to design and rig up but also to run. Still the belief was that it was possible to build a rational system in which the progress of education could be based on a variety of evaluations serving as grounds for central and decentralised political decisions as well as professional decisions.

---

[5] Franke-Wikberg, S., & Lundgren, U.P.: *Att värdera utbildning. Del 1. En introduktion till pedagogisk utvärdering.* [To appraise education. Vol. 1: An introduction to educational evaluation]. Stockholm: Wahlström & Widstrand, 1980 2:nd ed. 1980. 3:rd ed. 1985.

[6] Compare Granheim, M., Kogan, M., & Lundgren, U. P. (eds): *Evaluation as Policymaking. Introducing evaluation into a national decentralised educational system.* London: Jessica Kingsley Publishers, 1990.

Quality Assurance

But facing the millennium and a rapid change of the economic and political landscape the anxiety in a world represented by media gave little space to rational decisions. The claim was not for knowing more what education is about in order to prepare for the future, but to go back to a lost world. Evaluation lost its prefix and was more and more replaced by quality assurance.

The wonder with the word of quality is that it can embrace all kind of definitions. Basically there are three ways of understanding the concept of quality.

The very word stems from Latin *"qua litas,"* which means a holistic with its specific characteristics. In *The Oxford Guide to the English Language* quality is explained as *"degree or level of excellence; characteristic, something that is special in a person or thing."*[7] This definition includes value judgement.

Thus according to one definition quality is a value judgement, i.e. the relation between the subject and the object. Hence, President Johnson, President Conant, Mrs. Hull (Sara´s teacher) and Mr. Tykociner (the man next door) can all agree on the necessity of quality in education, but have quite different ideas of what it is. The second way of defining quality deals with quality as fulfilling given standards. The quality of a McDonald hamburger is that it tastes the same in Urbana and Stockholm. The third concept is the Aristotelian one defining quality as a relation between the various subjects about an object; thus something that develops with an enlightening discourse. President Johnson, President Conant, Mrs. Hull (Sara´s teacher) and Mr. Tykociner (the man next door) have to talk to each other and find a definition of quality which they all can agree on.

---

[7] *The Oxford Guide to the English Language.* London: Oxford University press. 1988.

Thus, evaluation is really at stake. We have in Sweden developed and are developing, as I see it, the most advanced system for evaluation. It is a variety of evaluation models and methods. It responds to quite different needs and will respond to still more different needs. Politicians on national and local level can be informed, parents, grandparents and students can be. Or in other words we have never known so much about our school system as we do today.

But choices are not made rationally and political decisions are not taken on basis of grounded evaluations. The results of evaluations are to be understood as news. In a world of anxiety and fear for the future, where quality ultimately is a question of economic values only bad results are good results.

I still believe in the necessity of having good grounds for decisions even if they are not used they will in the long run enrich an enlightened discourse about education. It is in such a conversation that quality can be found and defined.

Closing the circle, what is really at stake is that I cannot see an intellectual rethinking of dominating ideas about educational evaluation and the use of educational evaluation. A rethinking that I hope is at Stake.

We need to be served a fresh vegetable beef stew.

Illogical Teaching

James Raths
University of Delaware

I am honored to have been invited to speak to this distinguished audience on this sublime occasion. During my tenure in CIRCE, I came to appreciate and enjoy our "brown-bag" lunches where students, faculty, visitors, and CIRCE staff were given the opportunity to share developing ideas, early research proposals, or incisive issues in the field of evaluation. The purpose of the luncheons was not to convince others to believe certain hypotheses, or to "show-off" one's erudition (at least not always), but to seek clarification, input, rival explanations, and other forms of help in the intellectual arena. Of course, Bob played a key role in setting the tone for the luncheons and in contributing important comments, suggestions, and insights. In this role, his endeavor was the "stuff" of intellectual leadership, of which we were all greatly appreciative.

Let me use this forum as a "brown-bag" lunch. I would like to share my experience working on a committee of distinguished educators who are working to revise Bloom's *Taxonomy* (1956). The committee is led by David Krathwohl who, incidentally, worked at the University of Illinois and whose resignation from the faculty here led to the hiring of Bob Stake. Lorin Anderson, one of Bloom's students, is a co-chair of the committee along with Krathwohl, and there are a number of other distinguished committee members working on the project.[1] Without being too self-effacing, I will tell you that once convened, the committee members thought that it would be useful to have a "teacher educator" join the group to give the project a perspective beyond that of education psychology. It is my understanding that 10 or 12 people were invited to play the "teacher educator" role, all unavailable, before I was invited to the table. So, it can be said that the committee is almost entirely composed of distinguished scholars and one teacher educator who has been working hard to contribute to the group's work.

---

[1] Other committee members include Peter Airasian, Kathleen Cruickshank, Richard Mayer, and Paul Pintrich.

I am commenting on a work in progress (Krathwohl & Anderson, 2000). I don't speak for the committee, but as one of its members. With these qualifications, let me proceed first by saying something about what the committee is doing, and then discussing some interesting instructional issues that emanated from this effort.

The Revision: Bloom II

Bloom's students tell us that one of Ben's deepest regrets is that very few people ever read the *Taxonomy* (1956). Instead, they read, often in general methods texts or in measurement texts, a re-print of the brief six-level taxonomy table published as an appendix in the original version. The committee planned not to make that mistake--and they are including in the revised taxonomy a series of teaching vignettes demonstrating how an understanding of the Taxonomy (revised) and its application to planning instruction can be helpful. It is assumed that the vignettes will make the Taxonomy (revised) more readable. The committee solicited six teachers to write vignettes describing a teaching unit--with objectives, thick descriptions of contexts, and accounts of the methods of instruction and the assessments.

Second, the committee wanted to shape the classification scheme to reflect the advances in cognitive psychology since 1956. The original work, crafted in the heyday of behavioral psychology, eschewed terms such as "understanding" and "thinking" in part because they did not give reference to observables. The committee was willing to speak of "understanding" in this new version and it substituted "recall," as a psychological process, for the term "knowledge" as the first level of the revised taxonomy. To accommodate this change, the committee introduced a new dimension to the taxonomy--a knowledge dimension. In this dimension, the committee included declarative knowledge, conceptual knowledge, procedural knowledge, and metacognitive knowledge.

Third, the committee defined an objective as having two components--a VERB (designating the cognitive process) and a NOUN (stipulating the level of knowledge that was involved). So, the two dimensional taxonomy has cognitive processes as columns, level of knowledge as rows, and educational objectives classified into cells. Consider the following examples:

1. Students will recall the six steps of the scientific methods.

2. Students will apply the square root algorithm.

In these examples, the cognitive processes are "recall" and "apply." The levels of knowledge are declarative (six steps of the scientific method) and procedural (the square root algorithm).

Of course, there are many other changes included in the Taxonomy (revised). The ones I have chosen to highlight this morning inform the issues that I plan to raise in the next section.

Instructional Issues

There are two issues I would like to address. The first has to do with teachers' conflating objectives and activities and the second has to do with teachers' using activities drawn from the higher levels of the *Taxonomy* to advance lower level goals. These and other issues arose as we began to study, edit, and think about the vignettes we solicited from teachers. We did not intend to advance our vignettes as examples of superior teaching. To the contrary, we wanted to say that the examples we were including with the Taxonomy (revised) were representative of teaching found in current classrooms. It was our goal to demonstrate that the Taxonomy (revised) was useful in informing analysis of teaching by the teachers themselves and by others.

Conflating objectives and activities. Perhaps nothing seems more logical, especially to educational psychologists who are interested in evaluating the impact of teaching, than to assume teachers can differentiate their objectives from their classroom activities.

The logic of instruction and instructional planning, as seen by some evaluators, leads to strong feelings of impatience with teachers who plan lessons in terms of activities rather than in terms of objectives. Imagine this conversation between a teacher and an evaluator from the university:

Evaluator: What objectives are you addressing in class today?

Teacher: My students are holding a debate about the Constitution.

Evaluator: But a debate is an activity. I asked about the objective for the lesson.

Teacher: That's it. Our objective is to engage in a debate!

Back at the university, the evaluator would likely cluck, cluck, cluck about how teachers are so inept at instructional design that they can't distinguish between objectives and activities. In our early drafts of vignettes, our teachers frequently wrote out objectives that were more activities than objectives--at least from our view. We were confident that teachers could make the distinction the evaluators valued if they wished. It occurred to us to ask a better question: What are some explanations for why some teachers frame their objectives as activities? Here are some explanations that seem worthy of consideration:

The first explanation is that with the recent emphasis on performance objectives, teachers see performances as objectives--and the performances are in essence activities (Wiggins, 1993). So, teachers write as objectives, "to write a letter to Congress," or "to conduct an experiment," or "to give a demonstration of using perspective in a drawing." Are these activities or are they implicit objectives?

On one hand, if the lessons teachers teach address the performance tasks so that students are "taught" how to write an effective letter to Congress, or they are "taught" how to conduct an experiment, or they are "taught" how to give a compelling demonstration of perspective, then the activity is indeed an objective.

Another explanation for the conflation of activity and objective is that the activity, as a culminating task of a lesson or a unit, allows the teacher to assess students' progress toward the objectives of the unit. In these cases, perhaps giving an activity as an objective is simply shorthand for: "*(To assess my unit objectives, I ask students)* . . . to write a letter to Congress, or to conduct an experiment, or to give a demonstration of perspective." The words in parenthesis/ italics are unspoken. In this mode, while the response to the evaluator's question may be not directly on target, it does focus attention on the ways in which the objective will be evaluated.

A final view is that some teachers are convinced that there exist educative tasks, worthwhile assignments, that have value in their own right. Some experts have said that education comprises what is left after we have forgotten all the specifics we were taught in school. What do we remember about our school experiences? We are more likely to remember a trip to the zoo, our participation in a dramatic debate, or our working hard to prepare a presentation to the Science Fair than we are to recall inert knowledge taught in lessons more to the evaluator's liking. So, perhaps teachers see "objectives as activities" as a strategy for engaging students in worthwhile, educative, provocative experiences that are fraught with learning potential (Peters, 1967). In these cases, the activity *is* the objective.

Returning to the definition of objective advanced in the Taxonomy (revised), we can see that the definition doesn't help address this issue. In fact, it heightens it. Examples of higher level objectives in the current drafts of the Taxonomy (revised) include the following:

1. (For analysis). To write a short summary of historical events. (Chapter 5A: p. 24).

2. (For evaluation). To evaluate a solution (e.g., eliminate all grading) to a social problem. (E.g., the need to improve K-12 education). (Chapter 5A: p. 33).

In a sense, these objectives appear to be activities. Once students write a short summary of historical events, what have they learned? And after having evaluated a solution to a social problem, what have they learned?

As an aside, there is another difficulty apparent in these examples. The committee has advanced the definition of objective to be a sentence in the form of VERB, NOUN where the VERB is a cognitive process (write, evaluate) and the NOUN is knowledge. The definition of knowledge is stipulated by the committee to include declarative knowledge, procedural knowledge, conceptual knowledge, and meta-cognitive knowledge. Our formulation of knowledge doesn't take into account "historical events" as knowledge, or "a solution to a social problem" as knowledge. So our examples don't seem to capture the essence of our definition of knowledge. The committee is wrestling with this problem.

Higher level tasks; lower level objectives. Bloom and his colleagues (1956) stipulated six levels of objectives in the cognitive domain that were linked together in a taxonomic relationship. A taxonomic relationship in education implies that the accomplishment of an objective at a higher level requires attainment of objectives at the lower levels of the taxonomy. To comprehend, for example, a student needs to recall a number of things; to evaluate, a student must also recall facts, comprehend passages, apply procedures, analyze data, and synthesize reports. Evaluators and teacher educators often advocated that in good teaching, there should be a match between the objective and the activities designed to lead toward it. So, if the objective was to "recall," then the appropriate activity would be practice in giving recall--perhaps with flash cards, spelling bees, or other forms of drill. If the objective were at the "application" level, then the appropriate activity would ask students to apply ideas to new settings or in new contexts. In this instance, the activity would match the objective in terms of its cognitive level.

Good teachers, in some instances, seem to engage students in higher level tasks for the purposes of learning lower level goals (Sanders, 1966). If they would like students to recall aspects of Macbeth, they engage them in analysis tasks and evaluation tasks. If teachers want students to apply scientific principles, they engage students in synthesizing experiments or analyzing the experimental work of others.

On their face, these practices seem to represent a mismatch of goals and activities. These particular teaching strategies, however, seem to take advantage of the taxonomic nature of the cognitive levels that Bloom et al. described.

Students working at the higher levels are rehearsing in important contexts the lower level objectives. Based on our understanding of "time on task" and its relationship to learning, it seems likely that the more engaged students are in higher level tasks, the more likely they will be to master the lower level objectives. It is also likely that higher level tasks are intrinsically more interesting to students and to teachers.

This strategy also helps teachers avoid a conundrum of sorts. The higher the cognitive level of an objective, the more complex is its assessment. Assessing higher level objectives is problematic and poses challenges to teachers in making standards explicit; in sampling a domain of behavior; and in giving grades. The latter problem exists because of an "age-old" maxim that seems to define fairness in some schools and classrooms--teachers shouldn't test what they haven't taught. Sometimes, assessments of higher level objectives by necessity tap novel areas and call for some forms of transfer--a challenge that some students see as unfair.

Thus, teachers can work in the best of both worlds--engaging students in higher level tasks for the purposes of advancing lower level objectives and assessing at lower levels of cognitive challenge to avoid enduring problems of assessment.

Summary

This essay attempted to advance explanations for behavior that may seem to be illogical--confusing of objectives and activities and employing classroom activities that are mismatched with instructional objectives. Several tentative explanations were offered. Surely no single explanation accounts for the actions or decisions of any individual teacher, and there may be complex reasons for any given decision. We need to study teaching and teachers' thinking in more detail before we can have confidence that these explanations are credible.

References

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: Handbook I: The Cognitive Domain.* New York: Longman.

Krathwohl, D. R., & Anderson, L. W. (Eds). (2000). *A taxonomy of learning for teaching: A revision of Bloom's Taxonomy of Educational Objectives.* New York: Addison-Wesley-Longman.

Peters, R. S. (1967). *Ethics in education.* Oakland, NJ: Scott, Foresman.

Sanders, N. M. (1966). *Classroom questions: What kinds?* New York: Harper & Row.

Wiggins, G. P. (1993). *Assessing student performance.* San Francisco: Jossey-Bass

Love and Death and Responsive Evaluation

Saville Kushner
University of East Anglia

In this address I want to attempt three things. First, I want to recover a notion of "authenticity" from the ashes of the postmodernist passing and claim that evaluation is characterised by its capacity to break through social and political artifice and generate "more authentic" accounts of social life and citizen needs. Second, I want to bring into question our continued acceptance of "programme" as an appropriate focus for evaluation activities and seek to redirect our attention to the lives of young people. Finally, I will end by arguing that a primary application of a focus on young people might be to identify educational standards in collaboration with them, relevant to their lives and needs, as a strategy of resistance to the imposition of politically-driven standards in education.

Death

"The walls of society," wrote Peter Berger (1963), "are a Potemkin village erected in front of the abyss of being . . . a defence against terror." We are bound by our fear of mortality into social artifice, diversionary tactics--inauthentic roles, forms of organisation in flight from moral responsibility. The grandest artifice, of course, is the Hobbesian State. For Seery (1996) the failure to escape from a Hobbesian social contract founded upon the fear of death has tainted democracy and made of it "a second-best compromise, a calculated risk." The Hobbesian contract is the secular version of the religious exploitation of mortal fear on which is constructed unimpeachable authority--the outer limits of freedom. He condemns the absence of thought and debate about mortality for this is what prevents the emergence of more sophisticated--e.g. rights-based--versions of democracy.

For Berger, concerned with humanism more than specifically democracy, social enquiry cannot be so bound into the artifices of role and organisation. Social science exists to monitor the state of these social compacts and the extent of the fictionalising. Through engaging in the act of enquiry we face--we are obliged to face--the terrors--the closest we come to objective truth. We occupy a role, indeed another social construct, but one that is privileged by its search for authentic expression and by standing somewhat outside of normal social relations (postmodernist objections notwithstanding).

Death, in its corporeal, most mundane form, is urgent and real enough a theme for educators and educational evaluators. Read Linda McNeil's critique of an emerging US National Curriculum which starts with the words of a young boy saying school is a refuge from killing and being killed on the streets. Note: reality (the other word for death) is just another poverty disease: artifice, evasion and inauthenticity come easier for the middle classes. Dismayingly, what awaits that boy in school is hardly the kind of confrontation with those realities that will eventually allow him to cope with them. School is in the vanguard of the Hobbesian flight.

And, too, my interest in this theme was first sparked when I conducted a case study in a hospice for the terminally ill. There, my evaluation was limited by the fears and tolerances of those who lived and worked in the hospice. The Mother Superior, the Chaplain and the senior medics were all people who were touched by mortality and who transferred their fears--each in their own way--into forms of professional practice and forms of exchange with both patients and families. Where my questioning and my portrayals threatened to articulate those fears--just to give them form--my work was disciplined with recourse to our confidentiality contract and I became complicit with the avoidance strategies. Where I insisted on exposing the interaction between fear and action--publishing an account--there was an attempt at suppression.

Death (says Mellor, 1993) is a threat to the modernist project since it puts a limit on personal projects and, thereby, to our commitment to societal goals--it reduces the attractions of change. Reflections on mortality remind us of the incompleteness of all projects. Hence it is, as they say, privatised--hidden from view, outlawed--as, nowadays, are non-compliance, dissent, failure to meet targets and other sources of important learning. And so this of my two operatic themes (Love and Death) stands for less urgent possibilities. Death in the context of our educational concerns stands for incompleteness, failure essential for learning, intractable authority--the ever-receding and non-reachable standard.

Doctors in the hospice, for example, knew well that medics traditionally reject the death of patients as sign of their medical failures, of the limits to medical knowledge, and so marginalise it (and the terminal patient) from their professional lives. One key mission of the hospice movement is to recover confidence in medical practice--i.e. learn how to come to terms with the limitations of knowledge. Many politicians have barely started to address a similar condition in our education system.

Nor am I claiming that we need to take a lugubrious and negative view of what stands for death. Quite to the contrary--the insistence on inauthentic compliance to the policy plot is a kind of death in itself and a denial of life--i.e. a denial of diversity and idiosyncrasy. "All plots," writes Don DeLillo (in his book *Libra*), "lead to death." In its own way, the hospice accepted and promoted death (complete pain control allowed the Mother Superior to claim that a dying person was "the best audio-visual aid we've got")--the theory was that its acceptance brought a liberation which itself allowed a dogged celebration of life.

In education as in life, mortality is the key issue, death the main protagonist. If we were not haunted by the ephemeral nature of our accomplishments we would not, perhaps, be so obsessive about promoting them in schools. The situation is serious for youth who lie on the wrong side--albeit the fortunate side--of the most fundamental paradox in schooling. Here, for the most part, are people whose consciousness of mortality is barely ignited, but whose same consciousness is being tampered with by people for whom mortality is a never-simmering reality. Here is a hidden struggle, as portentous as it is unnoticed.

The sensitivity of this situation is intense--the danger of an accidental scuff creating an explosive spark in a young mind. I often hear artists in schools talking of wanting to "pass on the spark of creativity" to the child--as though creativity were an immortal and honorific blessing. A student of mine--an English teacher--talked to me of the personal pain of trying to teach Beckett to his pupils--how do you explain "Waiting for Godot" without contaminating that luxurious moment of immortality? But then I frequently recall a moment in one of my evaluations when a young (8-year-old) Muslim girl explained to me why, when she joined music workshops in schools, she risked inheriting a narrowing grave for her sin. Too late for the "spark" to do much more damage there.

I noted this last datum on the evaluation of an orchestral outreach programme, and the story raised a question about how we view educational programmes themselves in relation to those who people them. I asked another child, Richard, from the same school what it was like to be a pupil--"I don't know," he said, "I've never been a teacher." Well, intentionally or not Richard makes us think of how we lock children up in our educational Potemkin villages, intrigued more by the gravity of our campaigns than with the experience of living in inauthentic states; how we so consistently fail to measure the significance of that campaign in the immortal life of the child, but how obsessively we assume the place of that child in the significance of our ephemeral strategies. So I want to look at educational programmes we evaluate.

Of the existential tricks Berger counts among the Potemkin edifices the programme stands tall. Here is the bulwark against failure, the key vehicle in the modernist forward-moving convoy. Programmes, the mythology goes, once were the social scientists' long-yearned-for laboratories of change, the observed experiment writ-large, where social process could be dissected and analysed, bombarded and altered and then announced to a waiting world. Small wonder, and for good reason were evaluators attracted to them. Twenty years ago Carol Weiss wrote of the cooption of evaluators into programme realities and their being career-enmeshed with them. And so we are. One of the underlying biases we live with is our frequent assertion of programme status over that of the individual. Look at the contents page of almost any evaluation report. Context comes first, and that almost always means programme and policy contexts. Young people (where they appear) come later.

This would not be so calamitous if programmes were the speculative theatres of observation they once supposedly were. Now, however, they are unmistakably the purposeful "colonisers of the future," demanding loyalty to progress, intolerant of hesitancy in respect of change. They are the harbingers of Don Cambell's "experimenting society"--thoroughly imbued with the ideology of progress and scientific authority; saturated with inauthenticity and intolerant of failure and incompleteness. As I recently heard a radio broadcaster say, we live in a world where there is no longer a "Plan B."

Our tendency to "read" children's lives through the lens of the "school" or "curriculum programme"--to use the programme to shed meaning on the work and lives of so-called pupils--signifies further cooptation into "Plan A" and a flight from mortality and tolerance of failure. When evaluators believe in the social status of a social programme and use it as a template of meaning placed on individual thought and action--i.e. when evaluators go along with the artifice of role--this one a "teacher," this a "pupil," that one a "manager"--we, too, engage in evasive action and become part of the exhortatory machinery that drives people on. We need to come at programmes "from an angle."

Love

The alternative, of course, is to document people's lives and to use these as contexts in which to read the significance and the meaning of the Programme--i.e. to invert the relationship between programme and person. If I am hard-headed about anything it is this--that in educational evaluation almost all that is intrinsically worth researching are the lives and views of young people; most of all else is avoidance and cooptation. This means a key evaluation task is measuring the significance of programmes in the lives of young people--rather than the inverse of that--and, of course, documenting how educational programmes consistently (and importantly) fail them. And this means little more or less than talking to young people.

Here we walk in less familiar territory for it requires evaluators to engage in an immersion programme--immersed, that is, in young people's lives. But the point is to break the link between programme and progress--to search for Plan B--as often to frustrate and not to service decision making. We need, as one of my students once alleged of me, to be "in love" with our respondents.

This was a moment when I exposed my students to the questionable privilege of wading through (you might dignify this by saying "deconstruct") an archive of one of my evaluation projects which was located in a music conservatoire. I asked them to identify me and how I appeared in various guises. "It's obvious," said Ed, "you were in love with the students!" And so I was--though I have to say in a social-cerebral form of the affliction, which is how Ed meant it.

Well, I have written about this (Kushner, 1996) so I will not dwell too much on it here. What Ed did mean was that he noticed evidence of mutual dependence, mutual exploitation, joint celebration and a fascination with the emotional precipice of social intimacy. Here was evidence of engagement, an intermingling of interests--but, ultimately, as in all good tangos, of final betrayal. I talked as a friend but slunk off to write as a scientist—"the eyes of a sinner, the hands of a priest," as Sting's lyric goes.

The point about this is that this is what is involved in the privileged role hinted at by Peter Berger--the social enquirer who cannot enjoy the luxury of inauthenticity, who comes at our edifices to inauthentic experience from the angle of immediate perception. To document the lives of young people involves an essential betrayal--a drawing close and an eventual distancing.

Responsive evaluation

I started out on this track, actually, encouraged by Bob Stake's notion of portraying "the mood and even the mystery" of a programme--"mood" and "mystery"--two words I least expected to read when being inducted into programme evaluation. I still consider this to be a radical aspiration yet to be widely realised by us. Here--I suppose to love and death--is where this has led me for here lie programme mysteries. I do not lose my interest in programmes and nor my obligation to report on them. But I think we can do a more accurate job of measuring their significance than we do--we ought to do more of a job to locate programmes as iterative renewals of the social contract and to see each, thereby, as an opportunity to re-evaluate that contract and to expose its artifices. It is Thomas Hobbes, not John Stuart Mill, who hovers as the dark eminence over the field of evaluation.

So I worry about the continued focus on programmes in the Responsive approach. I worry that in treating the programnme as "stimulus" we are dealing with the surrogate, and that what we need to do to properly understand programmes is to forget about them for a while.

There is, in this respect, a particular application of my proposed inversion between programme and young person, and it relates to another of the monolithic artifices which looms menacingly over education--standards. The elegance of the myth, the sheer aesthetic neatness of the concept of a reachable standard renders it virtually unimpeachable in public discourse. Here is the hardest clause in the social contract between educational practitioner and citizen--achievement delivered in exchange for social status. We cannot, in my view, resist this movement fighting, as we have to, with the clumsy, Heath-Robinson weaponry of complexity.

What we might do, however, is to expose the artifice with the undeniable voice of the "client"--the young person (by which I include their families, of course). A key task for evaluators of educational programmes might to be to work with young people to identify what counts for them as reasonable and relevant educational standards. I am not talking of administering student "happy sheets," nor of chance interviews asking students' views of school. What I propose implies more complex methodological strategies. They are informed views we must seek, educational criteria discovered out of comprehensive analyses of lives, sociologies and school experiences. We need to approach young people not merely as the sources of information and data, but as participants in the process of analysing and understanding data.

This way, at least, we might generate accounts and visions of schooling suffused more with a celebration of life than with the submissive awareness of its passing.

References

Berger, P. (1963). *An Invitation to Sociology,* London: Penguin Books.

Kushner, S. I. (1996, Fall). To have and have not: Critical distance and emotional proximity in music education evaluation with a critique of "connoisseurship." *Bulletin of the Council for Research in Music Education, 130,* 52-64.

Mellor, P. H. (1993). Death in high modernity: The contemporary presence and absence of death. In D. Clark (Ed.), *The sociology of death.* Oxford: Blackwell.

Seery, J. E. (1996). *Political theory for mortals: Shades of justice, images of death.* London: Cornell University Press.

Stafford Hood

Susan Noffke

Rita O'Sullivan

James Sanders

Responsive Evaluation Amistad Style: Perspectives of One African American Evaluator[1]

Stafford Hood
Arizona State University

I must admit that I struggled to come up with a title for this presentation. I was torn as to whether I should keep the title and my remarks "light" by taking a few humorous jabs at Bob or to take a more serious approach. In thinking back, I have had a few interesting moments and conversations with him over the past 10 or 15 years while I was a student here and during my post U of I years as I tried to make progress and sense of the twist and turns in my professional and personal life.

Many of us can likely relate to--if we use Bob's words--a "shared experience" with him either once, a few times, or many times. This shared experience is that at one time or another, in a one-on-one conversation with Bob, or possibly in a group, he has been known to take you places during some very powerful verbal discourse on measurement, program evaluation, or the meaning of life and you would not know how or why you were there and more important if you wanted to be there.

As I thought about this particular occasion and my brief moments to speak to you, I decided that I would not waste my precious minutes in an effort to entertain you and /or roast you, Bob. So I have chosen to present my remarks, personal and biased as they may be, as they were inspired by the title, *Responsive Evaluation Amistad Style: Perspectives of One African American Evaluator.*

> In 1839, 53 Africans who had been kidnaped from Sierra Leone mutinied aboard a Portuguese slave ship, killing all but two of their captors. They ordered the men to turn the

---

[1] A revised version of this paper will be published in V. G. Thomas and C. Ellison (Eds.). Educational Equity and Excellence in the African American Community: Moving Beyond National Standards and Assessment.

> schooner around, but the two sailors duped them, heading to Africa by day and America by night. Two months later, the Africans were in a Connecticut jail, facing charges of piracy and murder (Schneider 1998).

Lewis Tappan, a Christian abolitionist, led his group in an effort to defend the Africans and hired lawyer Roger Sherman Baldwin. Baldwin would later be joined by John Quincy Adams in this legal struggle to free this group of Africans. For many John Quincy Adams is primarily known as the sixth U.S. President (1825-1829), the only President who was the son of a President (John Adams, 3rd U.S. President 1797-1801), the president who swam nude in the Potomac River every day, weather permitting, or "Old Man Eloquent." But for some of us it was his role in arguing the Amistad case before the Supreme Court and resulted in the Africans being set free.

Anna Marie Madison (1991) and others (Wilcox, 1984; Chevalier, Roark-Calnek, & Strahan 1982) have implied that a responsive evaluation approach is one of very few approaches that accepted culturally diverse factors as being central to an evaluation. As I thought further about the Amistad incident I wondered whether it could serve as a lense for me to better understand responsive evaluation and assist me and hopefully others in conducting culturally responsive evaluations. I do believe that some of us already hope and feel that we have been conducting such evaluations. However, I wonder whether we have aggressively sought to refine the methods we use in planning, collecting evaluative information, analyzing, interpreting, and making recommendations while conducting an evaluation that is truly culturally responsive.

One of the obvious similarities between the Amistad case and the evaluation of education programs is the participation of African Americans as experts in a professional endeavor that could decide the fate of the stakeholders of color. Unquestionably, the outcomes of the Amistad case extended beyond the group of Mende who were on trial. The fact that the initial charge of murder and mutiny were dropped by a lower court because it had occurred at sea on a vessel under the protection of the Spanish crown and the U.S. courts had no jurisdiction to impose punishment (Barber, 1840). However, the issue that would remain before the Appellate and Supreme Courts was whether the Africans were property to be returned to Spain even though the slave trade had been outlawed by Spain, the U.S. government, and Britain. The political pressure on President Martin Van Buren, by the southern states, to support a legal determination of the Amistad Africans as property had serious implications for the legal status of slavery in the U.S. and would hang in the balance until the Civil War 20 years later. Therefore, the implications of the case on the future of African Americans (free and slave) would suggest the essential participation of African Americans on the Amistad defense team. Of course their participation in this capacity could not be expected, it was 1839. The point is that there were no legally trained African Americans available and one could question if they were would they have be given the opportunity to participate. This example is pertinent for my concern regarding the limited number of trained African American evaluators and their participation in the evaluation of educational programs that serve African American students. To make this point I do not think it is necessary to provide you with the numbers, but rather ask you to rely on your personal recollections as evaluators. I simply ask you to remember the number of African American evaluators you have come in contact with at research and evaluation units in central school district offices, state departments of education, and the U.S. Department of Education. How many African American evaluators have you seen as members of external evaluation teams evaluating educational programs that target African American students or even directing such evaluations? My guess is that most of your experiences have been like mine and would result in answers to these questions being very few and I would not be surprised if some would say none. But the response would likely be followed by the comment "it has gotten better over the past few years."

I believe that few of us would disagree that one of the major reasons for this situation is that graduate programs with the capacity to train program evaluators have not done enough to rectify this situation. The most telling symptom is the dearth of doctoral degrees awarded to African Americans and other groups of color by programs with such capacity. My on-going monitoring of the IPEDS data of doctoral recipients by institution, race and program areas within education at major research universities support my observations (Hood and Freeman, 1995). And for those who are interested I can provide these data at a later time. Yet the other telling symptom is the absence of African Americans on the faculties of programs with the capacity to train a cadre of program evaluators of color.

As we know more faculty of color will attract more students of color. Their presence is more likely to be viewed as evidence of receptiveness to culturally diverse research interests and commitment to mentoring culturally diverse populations as students and presumably as professionals. These factors are important for recruitment, graduation, and professionalization. These same factors would be effective if we were serious about increasing the number and participation of program evaluators of color. My personal interest for more trained African American evaluators is what they can contribute to "understanding" in the evaluation of programs serving students from this population.

Responsive evaluation places a premium value on "understanding" because it "tries to respond to the natural ways in which people assimilate information and arrive at understanding" (Stake, 1972 and 1975). The assimilation of information for the purpose of understanding will be strongly influenced by the cultural experiences of the stakeholders. As I listened to Edmund Gordon's recent invited address, at the 1998 Annual Meeting of the American Educational Research Association, some of his comments seemed applicable for my continued thinking about program evaluation in general and responsive evaluation in particular. Even though his comments centered around the limitations of traditional scientific methods employed by social scientists as they attempt to derive meaning from the behavioral adaptations of diverse populations, his observations are germane to the practice of evaluation as well.

Gordon reiterated that the research community is first responsible for producing knowledge as clearly, as validly, and as objectively as possible and secondly to pursue understanding. The responsibilities are shared by the program evaluation community but with a slightly different twist. In program evaluation the production of clear, useful, and objective knowledge and the pursuit of understanding is for the purpose of determining worth. In this case I emphasize the importance of the evaluation resulting in an "understanding" of the program, its value for those who are intended to be served, and its refinements to improve the benefits. I would argue that an evaluator's understanding of a program as it functions in the context of culturally diverse groups is the most critical dimension for evaluating programs that serve these populations.

We must honestly assess whether in our evaluation practice, concerning diverse people, potentially important aspects of diversity and its implications have not been ignored. We must safeguard against producing evaluative knowledge "that seems counter intuitive to the [culturally diverse stakeholders] and seems to contribute little to our understanding of the people. . ."(Gordon) and the programs which intend to serve them.

Responsive evaluation relies heavily on interviews and observations to achieve stakeholders' understanding of the evaluand and its perceived value or worth from multiple stakeholders' perspectives. I agree with Stake in that "human observers are our best instruments [and] the evaluator should not rely only on his/her own powers of observation, judgment, and responding [but rather enlist] a platoon of students, teachers, and community leaders" (Stake, 1975). I would only add that an effort to insure that observers in an evaluation of programs serving culturally diverse populations should include evaluators and observers who share a "lived experience" with the cultural group. Gordon referred to the work of an anthropologist, Michael Jackson, who queried "whether the lived experience is a necessary condition for valid observations." It was his view that "there was a possibility of our inability to understand the experience of the other." In my opinion, central to the observation is **the meaning of what has been observed**.

Nonverbal behaviors are particularly pronounced among culturally diverse populations. One African American psychologist, Naim Akbar (1975 as cited in Hale-Benson 1982), describes a few of the nonverbal behaviors in African American children. He notes that the African American child "expresses herself or himself through considerable body language . . . adopts a systematic use of nuances of intonation and body language, such as eye movement and position . . . and is highly sensitive to others' nonverbal cues of communication." When observing African Americans participating in the program under evaluation much could be lost towards reaching "understanding." Too often the nonverbal behaviors are treated "as error variance" in the observation and ignored. The same can be true when interviewing an African American program participant and stakeholder.

Stake stresses in his 1975 discussion of responsive evaluation that "[a]n evaluation probably will not be useful if the evaluator does not know the interest and language of his audience." This knowing of the culturally diverse group's language in the collection, analysis, and interpretation of interview data for evaluative purposes also requires attention to cultural nuances in how the language is expressed and the meaning it may hold beyond the mere words. The interviewer in a culturally diverse context may need to serve as an interpreter for the evaluator who does not share a lived experience with the interviewee. Janice Hale Benson (1982) discussed this difficulty as described by Borneman (1959) and Akbar (1975). Borneman (1959) suggested a circular approach to language is a dominant feature of African American culture. He stated

> In language, the African tradition aims at circumlocution rather than at exact definition. The direct statement is considered crude and unimaginative, the veiling of all contents in ever-changing paraphrases is considered the criterion of intelligence and personality (as quoted by Benson 1982 p. 41).

Akbar (1975) similarly asserts that African Americans "[rely] on words that depend upon context for meaning and that have little meaning in themselves . . . [while also] . . . using expressions that have meaning connotations." Therefore the review of interview transcripts without the ability to interpret meaning based on these unwritten rules could possibly result in interpretations that are more frequently wrong than right, thereby, limiting communication and ultimately understanding between the African American participant/stakeholder and the evaluator. Another example from the Amistad case may provide further illumination of this challenge.

One of the major difficulties which faced the Amistad legal defense team was the language barrier between them and the Mende defendants. In order to present an adequate and compelling defense, the defense team and the court needed to hear and understand the Mende defendants' story of the incident. The first attempt by the defense team to find an interpreter failed. The assumption that any African could communicate with the Amistad captives was erroneous. One of the members of the defense team (James Leavitt) brought "an old African" who claimed to speak the Congo language to the defense team's initial visit with the Amistad captives (Martin 1986). African as he may have been, the home of the Mende was not the Congo but rather Sierra Leone. The desperate circumstances of this failed attempt resulted in Leavitt writing in this first report,

> with these unfortunate persons who have been committed to prison and bound over to be tried for their lives, without an opportunity to say a word for themselves and without a word communicated to them explanatory of their situation (Martin 1986 p.12).

Lewis Tappan was more successful as he solicited the help of a Yale linguist (Prof. Gibbs) and John Ferry. There is a conflicting account that John Ferry, who was reported in one source as white, had spent some time in Mendi and spoke the language also served as an interpreter.[2] But another accounted indicated that the Mende captives reported that they had never seen a white man in their homeland. Finally, two Africans were found on a British brig of war ship. One of them had been freed from a slave ship by a British naval vessel and was now a sailor on one of the British brig of war ships. This man had been raised in Mendi as a boy before he was captured to be a slave. But after being freed by the British naval vessel he was taught to read and write English and then assumed the name James Covey. James Covey served as an interpreter and because of his "lived experience" as a Mende he became a trusted friend of the Mende captives. His involvement was critical to the Mende's defense not only as an interpreter but also as their voice on the witness stand.

Covey was able to facilitate an understanding of not only the Amistad incident but also the two worlds which had crashed together. His ability to explain the court proceedings and the implications to the Mende allowed them to play a more active role in their defense and partially to bridge the language and cultural barriers that existed. However, even with the language barrier partially bridged an understanding between the two cultures was still difficult. One of the most poignant examples (portrayed in the movie Amistad) was after the Appellate Court ruled in favor of the Mende and that they should be returned to their homeland. After this ruling and influenced by President Van Buren the decision was appealed to the U.S. Supreme Court. When the Mende were informed that the case must be retried by another court such action was beyond their understanding, since their cultural experience of justice was once a decision had been rendered it was final. When the Mende inquired that since the Appellate Court had ruled in their favor would not this also be true at the Supreme Court. Baldwin's reply for translation by the interpreter was "maybe." The interpreter replied "the word or the concept of maybe does not exist in the Mende language." A couple of lessons can be learned from this excerpt of the Amistad story that may be relevant if we seriously try to extend responsive evaluation to culturally responsive evaluation.

---

[2] Following the presentation of these remarks at Robert Stake's Retirement Symposium a review of Barber 1840, Johnson 1990, and Martin 1986 corroborated that John Ferry was an African. Martin (1986) reported that John Ferry was from the Kissi tribe and "had been unable to speak enough Mendi to prove effective at the trial" (p.5).

First it is apparent that James Covey's role was more than one of interpreter. He was the portal between two conflicting cultures. He interviewed, interpreted, observed, and reported. He was a participant observer for both the Mende and the defense team. His lived experiences in both worlds made him essential to the case. He was the vehicle that made the defense culturally responsive and to the defense team's credit they knew that such a person was essential to their endeavors. I believe the same is true for responsive evaluation.

A second lesson is in the search for an interpreter. The "old African" who claimed to speak the Congo language shared race with the Mende but not language. For the sake of argument let's say John Ferry had been a white man. He would have been more credible interpreter because he had lived among the Mende and spoke their language. James Covey was the ideal interpreter but had he not been found, John Ferry would have been a viable alternative. Therefore, a culturally responsive evaluation approach could include evaluators, observers, or interviewers who do not share the racial background of the culturally diverse group of stakeholder/participants. However, the extent of their lived experience in the cultural context of the participants and understanding of the group's verbal and non verbal communication must be closely scrutinized. I believe that Stake's responsive evaluation approach could accommodate some of the lessons learned from the Amistad case. At the same time, I believe that these steps have the best chance of being implemented if we commit ourselves to increasing the number and participation of trained evaluators of color.

As surely as there were Amistad's in the 19th Century, there are psychometric pirates in the sea of educational evaluation in this century and probably await us in the next. They are not likely to hear the call I am making and will indeed question the value and relevance of what I have said today. I would expect this because my remarks could be viewed as agitation. Nevertheless, I am reminded of the words of Frederick Douglass, sixteen years after the Amistad decision. He wrote:

> Those who profess to favor freedom and yet deprecate agitation, They want rain without thunder and lightning, They want the ocean without the awful roar of its waters ... Power concedes nothing without a demand. It never did, and it never will (Douglass 1857, as quoted in Hale-Benson 1982).

More than a few of you have made contributions to what we do as researchers, educators, and evaluators. With this in mind it also became apparent that this may mark the beginning of some of you passing the torch to those of us who hope that our light will shine as brightly for the generation that will follow us. You are in the position to insure that we, as the next generation of researchers, educators, and evaluators, who are in the process of refining our craft, carry on the work you have begun and also extend it beyond even your imagination.

I spoke earlier about John Quincy Adams' role in arguing the Amistad case before the Supreme Court in 1841. At the age of 74 he refused to stand idly by when the prevailing winds of the time were prepared to impose an injustice upon a group of men who were drastically different from him and his kind. The content of his two day oration before the Supreme Court openly criticized President Van

Buren and Secretary of State Forsyth's readiness to deny the Mende justice in the rightful claim for freedom. He stated:

> The charge I make against the present Executive administration is that in all their proceedings relating to these unfortunate men, instead of that Justice, which they were bound not less than this honorable Court itself to observe, they have substituted Sympathy! Sympathy with one of the parties in this conflict of justice, and antipathy to the other. Sympathy with the white, antipathy to the black (Argument of John Quincy Adams before the U.S. Supreme Court 1841).

His position was not popular but necessary. This is typically the case.

As Ralph Tyler can be considered to be the George Washington of Program Evaluation, we may say that we are here to honor Bob Stake as the Thomas Jefferson of Program Evaluation. My hope is that somewhere someone will emerge as the John Quincy Adams of Program Evaluation. I hope that I am wrong but I doubt that I will see a John Quincy Adams step up in my life time. So I shall look to the Derrick Bells, Kweisi Mfumes, Maya Angelous, Fred Rodgers and James Andersons. Indeed, I shall immodestly look to myself as, in the final analysis, we all must.

References

Akbar, N. (1975, October). *Address Before the Black Child Development Institute Annual Meeting*. San Francisco.

Argument of John Quincy Adams Before the Supreme Court of the United States in the Case of the U.S., Appellants v. Cinque and Other Africans, Captured in the Schooner Amistad. New York: 1841.

Barber, J. W. (1840). *A History of the Amistad Captives: Being A Circumstantial Account of the Capture of the Spanish Schooner Amistad, by the Africans on Board; Their Voyage and Capture Near Long Island, New York, with Biographical Sketches of Each of the Surviving Africans Also An Account of the Trials Had on Their Case, Before the District and Circuit Courts of the United States for the District of Connecticut*. New Haven CT: Published by E. L. & J. W. Barber. Hitchcock & Stafford Printers.

Chevalier, Z. W., Roark-Calnek, S., & Strahan, D. B. (1982, March 22). *Responsive evaluation of an Indian Heritage Studies program: Analyzing boundary definition in a suburban school context*. Paper presented at the annual meeting of the American Educational Research Association, New York City.

Douglass, F. (1857). *Excerpt from a Speech on West India Emancipation*, delivered at Canandaigu (in Quarles 1969, p. 354) quoted in Hale-Benson 1982.

Gordon, E. W. (1998, April 13). *Producing knowledge and pursuing understanding: Reflections on a career of such effort*. AERA Invited Distinguish Lectureship. Presented at the annual meeting of the American Educational Research Association, San Diego.

Hale-Benson, J. (1982). *Black children: Their roots, culture, and learning styles* (Rev. ed.). Baltimore: Johns Hopkins University Press.

Hood, S. & Freeman, D. J. (1995, Fall). Where do students of color earn doctorates in education?: The top "25" colleges and schools of education. *Journal of Negro Education, 64*(4), 423-36.

Johnson, C. H. (1990, Spring). The Amistad case and its consequences in U.S. history. *Journal of the New Haven Colony Historical Society,* 3-22.

Madison, A. M. (1991, Spring). Introduction. *New Directions in Program Evaluation,* 53, p. 1.

Martin, B. E. (1986). *All we want is make us free: La Amistad and the reform abolitionists.* Lanham, MD: University Press of America.

Schneider, A. (1998, January 9). Advising Spielberg: Tulane historian has spent his career studying the Amistad Rebellion. Academe Today. *Chronicle of Higher Education.* Chronicle of Higher Education Website (**http://www.chronicle.com).**

Stake, R. E. (1972). *Responsive evaluation.* ERIC Document ED 075-187.

Stake, R. E. (1973, October). *Program evaluation, particularly responsive evaluation.* Paper presented at a conference on New Trends in Evaluation, Gotenborg, Sweden.

Wilcox, T. (1984). *Evaluating programs for native students: A responsive strategy.* Paper presented at the Mokakit Indian Education Research Association's International Conference, London, Ontario, Canada July 25-27, 1984.

Who Knows?, and Other Questions I Might Ask Bob Stake

Susan E. Noffke
University of Illinois-Urbana/Champaign

Good Morning!

The Backdrop

I thought I'd start with a short story about my first opportunity to assess, evaluate and know Bob Stake. It was about 5 and a half years ago, when I was at the University at Buffalo, during the Ethnography in Urban Education Research Forum in Philadelphia. My husband had joined us in Buffalo a few months earlier, after "commuting" (very strange term for a very unsuburban phenomenon) for three years from Madison. But he'd also accepted a position for the following fall at Illinois. We were all (kids included) hoping that I, too, would be able to find a job here. We had a lot "at Stake."

I did a session at the Forum with my colleague, and later co-editor of a book on action research (Noffke & Stevenson, 1995), on "The role of data in action research." In the session I said some of my usual stuff about data not really existing apart from the social relationships that construct them as evidence within particular groups and for particular political agendas. One gentleman at the session seemed quite intent, even distraught by what I said, and asked a number of short, but on target questions. I didn't think my responses satisfied him much.

After the session was over, my colleague asked if I knew who that man was. I didn't. It was Bob Stake. Confident that I had just ruined my family's life with my rather unusual if deeply held thoughts, I went to the reception. Stake was there. Lesson one, about knowing: Believe deeply in what you say and write. Who knows who's listening and reading? There's a lot at Stake.

Actually, I found the conversation really enjoyable (I don't know what Bob thought). It was wonderful to have someone listen so careful and talk almost as slowly as I do.

Lesson two, about assessing and evaluating: Listen carefully, and hear more than you say; always try to learn, not to prove. There's a lot at Stake.

As I recall, at some point after I came for my official interview, I sent him something of mine on the history of action research to read. He sent it back with useful and insightful comments. He supported me in coming here and has remained a very important person in my life here, even attending our son's recital last year, holding my hand while I played each note with Andrew and kept Laura quietly cuddled while her big brother played. When the time came to put the research part of my tenure papers together, there was no question whose views I valued. He was solid, but asked the simplest, hardest questions about what I was doing with my scholarly life I had encountered in a long time. I hope to do a bit of the same today.

The Paper

I approach this instance of thinking publicly about issues in assessing, evaluating, and knowing–subjecting my ideas to the public forum for a "validity" check, by using the same principles which guide my teaching. Betraying my long years as a teacher of elementary and middle school-aged children, a bit of butcher paper and crayons (or even markers–the bold "magic" of my childhood) are often used in classes to collectively take on the representation and discussion of ideas. Most needed in a graduate seminar, we often consciously "level the playing field" by charging small groups of students with the task of "representing" discussions of lofty concepts through this medium. It is an act of collective synthesis which, for me, reduces the privilege of those most comfortable with academic discourse and allows those most closely aligned with the lives of children–especially young children, a familiar medium. While I didn't bring my crayons today, I do see this short paper as my piece of butcher paper to share.

In our classes, I often remind myself that insofar as research is concerned I understand three simple and somewhat impertinent questions to be most important. For me, these questions serve as reminders that regardless of how elegantly or simply we address them, it is *discussion* of issues surrounding assessment, evaluating, and knowing in research that is of most importance. In all three questions, there is therefore a constant sense of contradiction. Through the asking of "WHO knows?," we come to recognize through the practice of research the integral ways in which our identities shape the boundaries of what can be known at the same time as we seek to open up the possibility of understanding things beyond them. These boundaries are reconstructed as we collectively and personally find spaces for action.

Who cares? For me at least, any attempt to construct a means of sharing or a means of evaluating what is shared (which is of course, what validity and reporting issues are about) begins by making clear both the values surrounding the research focus as well as those of the people whom it most clearly and deeply affects. It is about "whose knowledge?" but also about the meaning of caring--of interests and of interest groups. Whose issue is this? What meaning does it have to the daily lives and larger social, political, and economic contexts of those who live in a "practice"? How are the interests of the researcher(s) seen in relation to those of others connected to the practice being studied?

How do you know? I have spent most of my adult life with children. I often wonder at the almost simultaneous claims to understanding things as they are and to a deep wondering about what is that is so often a part of children's thinking. The question of how we know raises issues not only of the process by which we claim to know something and the kinds of things we accept as evidence, but also the ways in which our identities and experiences shape those things which we believe we understand as well as those things which are not visible to us. In research that is deeply embedded in practice, there can be no simple reliance on methods of analysis deemed to be objective and neutral, or even subjective and interpretive. The very processes of data collection and analysis shape collective understandings and can form the basis for new forms of social solidarity: Knowing is in relationships to and with others involved in practice. In order to engage in research, there needs to be a recognition of the limits of our understandings--the fragility of our knowledge claims--as we engage in social practices which push at their boundaries. It is both how we know, as well as how is it that we do not know (and perhaps cannot know) that is at stake. How do we come to recognize (as educators confront daily in their practice) things we have not known or other ways of knowing? How is it that we have not heard, seen, or recognized them? Through such questionings, the effort is not to establish the known, but to identify the nature and limits of current understandings in order to engage in meaningful action. Which leads to the third question.

So what? This question impels us not only to name and justify the interests which have led us to our study and the things we have learned by engaging in practice and the study of practice, but to also identify actions the ways in which the contradictions we uncover help to shape actions which are "ethically defensible and politically strategic" (Please excuse my quoting myself!). Although it seems too obvious to mention, research is equally about knowing and about doing. I have gained much over the years both from memories of my interactions with children and work with others who struggle in and for teaching and from my interactions with people who have helped me to see "through a glass darkly" where I am in society. In my daily work as an educator, I constantly make decisions about ethics and politics in relation to my actions. A cluster of such decisions surround issues of how I choose to make my work "public" (invoking Stenhouse, 1983, p. 185, here). I hear, see, and feel at a concert, as I watch a dance; I am part of the "testimony" at a church group meeting; I witness the creation of a quilt signifying people's experiences; I learn with people participating in a slide-tape presentation of their research "findings." These events of "reporting" send me forward into new understandings of "so what?" But they also push me to question what the question means, not only in terms of knowing or even knowing "what is to be done?," but also in terms of thinking through what IS being done that might inform my/our practice. What is being done that my privileged positions (not only in the academy) have not allowed me to see? With whom does my mode of representation or reporting allow me to connect? What values/interests are evident in mode of representation? In what ways does my method of reporting signifying of a particular, implied "audience?" As my practice involves education, I return always to the question of "How does this make the lives of children and those who share their lives in and out of schools better?" (Asked with thanks to John St. Julien, who reminded me at a key point of this question).

One final cluster of "worries" I have about assessing, evaluating, and knowing in research: I have both worried about and hoped for a long time about the increased embedding of action research in the academy and in school staff development programs. Why/Is there a need for "distinctive and stringent criteria?" While I have been part of and continue to struggle for the legitimacy of action research in both of these contexts, I have done so with the understanding that the political economy of knowledge production is also the production of legitimation. Universities, departments of education, and school districts and indeed each of us as practitioners seeks to understand, but we also, whether overtly, under the guise of objectivity or tacitly, seek to legitimate ourselves. We can justify these through positioning ourselves in relation to oppressive social conditions. We can also recognize that we assess, evaluate, and know as much through what we DO as through what we know--and how we see the two as intertwined. We speak these messages of knowing--testifying through our lives and those of the children, students, parents, and community members who share our practice. We do not "give voice," but instead are part of the process of removing barriers for speakers and listeners, writers and readers.

The means of assessing, evaluating, and knowing cannot then be separated from our agendas as social actors--we come to know ourselves and those parts of ourselves which are built on the oppression of others. In so doing--as a result of that doing, we open up or "subject" ourselves to the scrutiny of others, always knowing that the power differentials are not equal. We create "representations" of ourselves, of "where we are at"--people-ing the forces that others feel and see, through aesthetic, spiritual, economic and political lenses. We see ourselves, in all our absences and preserved privileges. Both are aspects of human diversity in terms of power, and are related to the doing and reporting of research.

Through our research work, we hope, not for "validation" through our public sharings of our work, although warmth and solidarity do sustain us. But we mostly hope for help in understanding the contradictions, the consonances and dissonances in our "reporting," that will help us and others see spaces for the creation of new action and thought. I've set up this contribution to the panel not as a revocation of various theoretical resources. Indeed, issues of assessing, evaluating, and knowing can be usefully informed by a number of theories both from within research efforts, including those shaped by newer qualitative, feminist, and critical race theory. But it must also always return to the essential questions by practitioners involved in trying to understand their social world and also be informed directly through their theories and actions.

References

Noffke, S. E. (1995). Action research and democratic schooling: Problematics and potentials. In S. E. Noffke & R. Stevenson (Eds.), *Educational action research: Becoming practically critical* (pp. 1-10). New York: Teachers College Press.

Noffke, S. E., & Stevenson, R. (Eds.). (1995). *Educational action research: Becoming practically critical.* New York: Teachers College Press.

Stenhouse, L. (1983). *Authority, education and emancipation.* London: Heinemann Educational Books.

From Responsive to Collaborative Evaluation

Rita G. O'Sullivan[1]
University of North Carolina at Greensboro

Introduction

This paper traces the author's initial use of Bob Stake's responsive evaluation approach (1983) along a 15-year path that has led to collaborative/participatory evaluation. Along the way, Stake and his work have sustained and enriched the author's evaluation practice. Other evaluators also have contributed to the process. This paper also shares preliminary empirical evidence that supports the value of collaborative evaluation and demonstrates how such an approach can improve evaluation practice.

In The Beginning . . .

In 1983, I was faced with the need to complete the evaluation of a three-year program for teen mothers in the Caribbean. I inherited a massive dataset that had been compiled for 151 participating teen mother and 35 controls. The person who initially designed the evaluation, set into motion an evaluation that required the full-time commitment of two host-country project staff who, over three years, completed and coded seven separate interview protocols. After a month of keypunching my way through the coded data (this was 1983 remember), I found to my chagrin that while second pregnancy data were available for 85% of the participants, only 56% of the control group had continued in the study. I had 36 of 151 participants who had become pregnant for the second time but no way of knowing how this figure reflected on the program.

I also had other concerns about the evaluation design. At the time, the sponsoring government ministry very much wanted to know the extent to which this program had been

---

[1] Teaching to Diversity, School of Education, University of North Carolina at Greensboro, Greensboro, NC 27412.
336/334-4095 (phone & fax) RGOSULLI@ UNCG.EDU

effective. The pilot period was ending and the government was seriously considering assuming responsibility for continuation of the project. The U.S. sponsor, for whom I worked, wanted to know how participants, their parents, and community members perceived the program. In the design of the evaluation and development of the seven interview protocols, even though the external evaluator did take time to include questions about parenting that would further her personal research interests, she gave no thought to collecting information from the various project stakeholders about the assessment of the program.

My third concern was personal. I needed a dissertation topic. In a confluence of events, I was able to redesign the evaluation, complete it, and then use it as a case study for my dissertation. Enter Bob Stake, although little did I know then that what I thought of as a case study wasn't really a case study according to Stake (1978).

During my doctoral work, I had taken a course in educational program evaluation and was familiar with the various approaches that were then popular in the emerging discipline. My experience with evaluation, both internationally and in the United States, created a context by which to weigh the information that was presented in the doctoral course. Among the evaluation approaches presented, I had gravitated most toward Stake's responsive model as the one that best mirrored my beliefs about evaluation and what it might accomplish:

> I have made the point that there are many different ways to evaluate educational programs. No one way is the right way. Some highly recommended evaluation procedures do not yield a full description nor a view of the merit and shortcoming of the program being evaluated. Some procedures ignore the pervasive questions that should be raised whenever educational programs are evaluated . . .
>
> Some evaluation procedures are insensitive to the uniqueness of the local conditions. Some are insensitive to the quality of the learning climate provided. Each way of evaluation leaves some things de-emphasized. . . .
>
> I prefer to work with evaluation designs that perform a service. I expect the evaluation study to be useful to

> specific persons. An evaluation probably will not be useful if the evaluator does not know the interests of his audiences. During the evaluation study, a substantial amount of time may be spent learning about the information needs of the persons for whom the evaluation is being done. The evaluator should have a good sense of whom he is working for and their concerns . . . .
>
> To be of service and to emphasize evaluation issues that are important for each particular program, I recommend the *responsive evaluation* approach. It is an approach that sacrifices some precision in measurement, hopefully to increase the usefulness of the findings to person in and around the program. . . .
>
> Responsive evaluations require planning and structure; but they rely little on formal statements and abstract representations, e.g., flow charts, test scores. Statements of objectives, hypotheses, test batteries, and teaching syllabi are, of course, given primary attention if they are primary components of the instructional program. Then they are treated not as the basis for the evaluation plan but as components of the instructional plan. These components are to be evaluated just as other components are. The proper amount of structure for responsive evaluation depends on the program and persons involved (Stake, 1983, 291-292).

I used House's (1978) framework of eight evaluation models to set the stage for the logic of Stake's (1983) responsive evaluation. I argued that the TEFLEP external evaluator had narrowly equated evaluation with the behavioral objectives approach, and thereby, had ignored important decision making and transactional components required for the evaluation. I found support for this argument in Guba and Lincoln's *Effective Evaluation* (1981). Guba and Lincoln acknowledged that their work had been influenced by Stake's, and although they strongly promoted qualitative approaches in naturalistic settings as best suited to the evaluation of education programs, they allowed that: "There are times, however, when the issues and concerns voiced by audiences require information that is best generated by more conventional methods, especially quantitative methods" (p.36).

I redesigned the TEFLEP evaluation, expanding it to include interviews with relevant stakeholders: advisory council members, TEFLEP staff, ministry coordinators, community representative, participants, and their parents. This redesign provided the information the sponsoring agencies needed to make decisions about program expansion and participants' satisfaction. I solved the second pregnancy measurement dilemma by identifying an equivalent cohort of teens on the island who had delivered their first babies a year before the TEFLEP program began and were therefore ineligible for participation. The retrospective sample allowed me to report that the second pregnancy rate of 24% among TEFLEP participants compared very favorably to the 48% second pregnancy rate among the comparison group for an equivalent three-year period. Thus, Bob Stake's work provided a framework for my dissertation and support for my evaluation practice.

In The Middle . . .

In 1985, I began working at the University of North Carolina at Greensboro (UNCG) as a visiting assistant professor in the educational research area. I was hired to teach the graduate educational program evaluation course and some of the introductory educational research courses. In 1986, Dick Jaeger and I were grappling with a program evaluation design and Dick suggested that we invite Bob Stake, Ernie House, and Kathryn Hecht to collaborate. I had, of course, heard Stake speak at professional meetings, but was delighted at the prospect of actually working with him and getting to know him. In the course of collaboratively developing a modular evaluation design with the group (Jaeger, O'Sullivan, Hecht, House, & Stake, 1986), I added new ideas and practices to my evaluation toolkit. I also discovered that the real Bob Stake had more dimensions than the Stake whose work had informed my understanding of evaluation and dissertation. I was most struck by his insistence on making the components of the proposed evaluation meaningful to the clients. To do this he designed a graphic that demonstrated how each of the evaluation modules fit within the context of the program. It's not something I would have thought to do. It measurably strengthened the evaluation design and imprinted for me the importance of client understanding in the evaluation process.

During the next few years, now as an assistant professor of educational research at UNCG, contact with Bob continued. He came to do a short course on case study methods and to a May 12th group evaluation meeting at UNCG. The following year, he invited me to a May 12th group meeting that he hosted at CIRCE which focused on issues surrounding classroom assessment. Through those contacts, my understanding of evaluation expanded and matured. I developed a deeper appreciation for the importance of qualitative methods in general and their particular importance to evaluators who care to be responsive to clients' needs. The inquiry into classroom assessment caused me to remember the important role that evaluators play in questioning topical educational policy areas beyond our clients' immediate intents; evaluators need to be responsive the public's needs as well.

In 1990, I had a research leave from UNCG for a semester and to my delight it coincided with Bob and Bernadine Stake spending a semester at UNCG. Bob was slated to teach his course in case study methods, and I had the luxury and pleasure of participating as a student. Had I been able to travel during my research leave, one of my first thoughts would have been to go to the University of Illinois and study with Bob. As events unfolded, I expanded my skills while Bob and Bernadine also came to know my family better. The case study class was a learning experience from a variety of perspectives. Although I had had Bob's short course in case study methods, the semester-long contact appreciably advanced my understanding of qualitative research methods in general and case study methods in particular. Extended contact with Bob and Bernadine proved to be the ideal research leave for me.

Within the next year, I developed and introduced a course in case study methods at UNCG. Within our educational research area at UNCG, the only course where students encountered qualitative methods was in educational program evaluation. Students sorely lacked the training they needed to use the qualitative methods that interested them. By default, I had become the informal qualitative methods person in the department. Luckily, an undergraduate degree in anthropology supported this designation along with years of experience using qualitative methods in evaluations. The case study course with Bob bolstered my knowledge and my confidence.

In terms of my evaluation practice, the strengthening of qualitative skills was accompanied by a deeper appreciation of responsive evaluation. Responsiveness was not just listening to a program's evaluation needs but also anticipating the audiences' levels of evaluation expertise and depicting the results in ways that enhanced their understanding. This also often meant that the evaluator's job was to reflect the program back to the audiences in intelligible ways; the audiences could decide about the merit. There was much merit in naturalistic generalizations (Stake, 1978).

Beyond The Middle . . .

The expansion of responsive evaluation to include audience understanding of evaluation findings has led me for the past six years to focus on collaborative evaluation. Since the term is often used interchangeably with participatory and/or empowerment evaluation (the topical interest group in the American Evaluation Association is called Empowerment/Participatory/Collaborative Evaluation), let me define my intent. I prefer the term collaborative because it implies that people share responsibility and decision making. When a stakeholder is asked to provide information for an evaluation, technically they are participating in that evaluation, but they are not necessarily collaborators in the evaluation design. Similarly, program participants are usually not program collaborators in determining the content or direction of the program. I, therefore, prefer the term collaborative evaluation rather than participatory. My intent is that, to the extent that they are able, that program staff and other stakeholders should be considered part of the evaluation team. This does not relieve the evaluator of the overall responsibility for conducting the evaluation or producing evaluation results. My assumption is that evaluators are engaged because of the expertise they bring to the endeavor, and that leadership for the evaluation resides in that role. I believe that collaborative evaluation is empowering to participants. As such, it is a valuable positive outcome of the process but not an intended one as described by Fetterman (1996).

I view the collaborative evaluation approach that I use as a natural progression from responsive evaluation. Not only does the evaluation need to be responsive to the programs needs, but it also should be responsive to the needs of the stakeholders to find the evaluation useful and the needs of the community to have people informed. Thus, evaluators can improve the general state of evaluation by taking every opportunity to enhance clients' ability to appreciate, understand, and conduct evaluations. This is not just conceptually sound but practically useful as well.

Utilization of evaluation findings continues to be a central problem in the field (Ciarlo, 1981; Patton, 1986; Smith, 1988; Stevens & Dial, 1994; Weiss, 1971). Patton (1997) would probably argue it is *the* problem in the field. Some charge the evaluator with the responsibility for promoting evaluation use (Chelimsky, 1986; Cousins, Donohue, & Bloom, 1996; Knott, 1988; Mowbray, 1988). Along with others (Fetterman, 1996; Greene, 1987; Guba & Lincoln, 1989; Levin, 1996; Patton, 1988; Linney & Wandersman, 1996), I believe that involving stakeholders in the evaluation process will improve evaluation utilization. In part, program staff ignore evaluation findings, because they do not understand them or have not been involved directly in the planning and implementation of the evaluation process. Distanced evaluators, conducting distanced evaluations, fail to engage program stakeholders in the evaluation and thereby limit the potential for the findings to positively influence the program. Logically, if program staff are collaboratively involved in the evaluation, then their use and understanding of the findings should increase.

I am well aware of the debate in the field about appropriateness of evaluators' roles (O'Sullivan, 1995). More, recently I have considered Scriven's (1996a; 1996b) objections to collaborative evaluation and the potential co-optation of the evaluator, as familiarity with programs and program staff increases. Yet usually the advantages gained in program awareness, staff cooperation, access to information, quality of information gathered, and enhanced receptivity to findings far outweighs the potential for (not the presence of) biased findings.

As a direct outgrowth of my belief in the strength of the responsive evaluation approach, I have opted for collaborative evaluation. How this translates into my practice is that I design evaluations that engage clients in the evaluation. The level of engagement varies by program evaluation purpose and client, but generally I seek evaluations where clients want to collaborate in the process. I also find, in light of limited evaluation funds, that when clients are collaborators in the process, more thorough evaluation is possible.

An Example of Collaborative Evaluation

The collaborative evaluation approach that I use is best exemplified by the evaluation of a county-wide, comprehensive, early-childhood program that we have led for the past three years. The program has received about $6,000,000 annually from the state to support programs that assist families with children under six years of age so that all children in the county are ready for school success. With that aim, the program contracts with about 40 local agencies in the county to provide approximately 50 different support services in the general areas of: Education and Quality Care, Family Support, Health, Translation, and Transportation. The evaluation budget for this program has been about $40,000 annually.

The program director and a committee member from the evaluation advisory group visited me to discuss the possibilities for evaluation. They were in the first 18 months of operation and only six months into their first implementation year. The program was, and still is, politically sensitive in the state which meant that its existence could, in fact, be influenced by evaluation results. The evaluation challenges were impressive: the large number of agencies collaborating to provide services; the large number of programs; the limited evaluation funds; the political sensitivity of the program to evaluation findings. The fact that the services to be provided would vary greatly by individuals added to these challenges. A child might receive vision screening and no other services from the program; another child might receive subsidized day care in a preschool center that was working on quality enhancement supported by the program and their parents might receive home visits from another of program's support services.

This evaluation dilemma was similar in scope to the one I had faced 12 years earlier, evaluating the teen pregnancy prevention program in the Caribbean. In order to be responsive to the program's evaluation needs, this time I needed to use collaborative evaluation. Clearly, given the size of the program and the available resources to conduct the evaluation, the program contractors would have to become active participants in the evaluation. They would have to supply basic information about program services that they had provided to include with the state's mandatory quarterly reporting requirements. Beyond that, these contractors also would have to collect evaluative evidence about their program accomplishments (outcomes). The external evaluation team would need to spend time working with the contractors, set up data collection systems, and might be able to conduct a few focused studies on important evaluative outcomes (e.g., client satisfaction, quality care, parent education, etc.). The key to the success of the evaluation rested with the ability of the evaluators to engage contractors in this collaboratively evaluation process.

Luckily, I had been working on such a process (O'Sullivan & O'Sullivan, 1998) and could propose it to the program. Convincing evidence from the field had pointed toward the development of an evaluation approach that strengthened evaluation expertise from within programs to improve the likelihood that evaluation would be well utilized. The approach also had to consider common misgivings about evaluation among program staff and limited availability of program resources for evaluation. Evaluation Voices was developed to improve evaluation expertise among program staff using an innovative cluster networking context. Programs were clustered by interest area, so that contractors with similar program could share evaluation strategies, instruments, and concerns. This context was structured so that the participants would reconceptualize evaluation as a dynamic process that required their active participation and included peer learning.

We proposed using Evaluation Voices cluster networking activities as the way to begin assessing and strengthening evaluation expertise among the program's contractors. We held evaluation cluster meetings in the first year of the evaluation to orient contractors to evaluation, share the evaluation plan, explain state reporting requirements, help them draft annual evaluation plans, and share data gathering strategies. During these meetings and subsequent individual technical assistance visits we emphasized the importance of finding out what they wanted to know about their contracted activities which almost always coincided with what the program wanted to know overall.

The level of evaluation expertise varied greatly by contractor. A few programs were fairly sophisticated in their evaluation practice, while a corresponding number had really never collected service statistics before. Most were struggling through the first year of program implementation with the usual delays in hiring, opening new facilities, launching new programs, etc. The state added to these first year difficulties as it worked through its own program start up complexities which included changing the format of their quarterly reports three times. The first year's evaluation report (O'Sullivan, Clinton, Schmidt-Davis, & Wall, 1996) provided overall service statistics from programs, shared success stories, reported the results of a survey to identify quality care standards in the county, and began sharing information about county-wide indicators of importance (e.g., infant mortality, number of day care slots in the community, collaboration, etc.).

Building on the year-one activities, we began the second year of the evaluation by transferring the compilation of service statistics to the program office and working to strengthen contractors' evaluation plans. Evaluation Voices cluster networking meetings continued as the way this strategy was implemented. Contractors participated in cluster workshops on evaluation planning that was followed by individual technical assistance as required. During these workshops contractors were told that they would be asked to share interim evaluation results at an "Evaluation Fair" to be held mid-year. During the Evaluation Fair contractors were expected to report their results by clusters to their peers. At the same time, they were asked to submit a written report of mid-year accomplishments. The external evaluation team members were available to assist contractors with implementation of their evaluation plans. The external evaluation team also worked with the overall program to develop parent education measures, assess collaboration, and continued to report on important outcome measures. The Evaluation Fair was held and interim results summarized. At the end of the second year, interim evaluation results were updated and included as part of the second evaluation report (O'Sullivan, D'Agostino, Prohm, Roche, & Schmidt-Davis, 1997).

By the third year of the evaluation, the evaluation processes established during the first two years took root and successful patterns continued. Evaluation planning occurred during the beginning of the year, with the Evaluation Fair scheduled once again for mid-year. Demand for external evaluation services was such that evaluation team members spent 10-15 hours each week at the program office, providing technical assistance to contractors and staff. Most contractors saw the external evaluators as collaborators and requests for technical assistance increased. Not surprisingly, the quality of evaluation plans improved as did the timeliness with which they were submitted. External evaluation team members also were asked to assist with data analysis for contractor or program collected data. Additional work continued on the identification of parent education measures and other common instruments.

Most importantly, the quality of the evaluation findings presented at the Evaluation Fair improved dramatically. The details of these improvements are chronicled elsewhere (see O'Sullivan & D'Agostino, 1998), but the importance of these findings is extremely relevant to the discussion at hand. The move toward collaborative evaluation was justified based on the assumption that such an evaluation approach would measurably improve the quality and utilization of evaluation. The empirical evidence collected, while still preliminary, strongly supports the quality improvement supposition of collaborative evaluation. Plans to test the assumption that collaborative evaluation improves utilization are underway.

In Appreciation . . .

Tracing the past 15 years of my evaluation practice points to the consistent and considerable contributions by Robert E. Stake. I am grateful for the guidance and most appreciative of the assistance. I look forward to continued collaboration.

References

Chelimsky, E. (1986). What have we learned about the politics of program evaluation. *Evaluation Practice, 8*(1), 5-22.

Cousins, J. B., Donohue, J. J., & Bloom, G. A. (1996). Collaborative evaluation in North America: Evaluators' self-reported opinions, practices, and consequences. *Evaluation Practice, 17*(3), 207-226.

Ciarlo, J. A. (Ed.) (1981). *Utilizing evaluation.* Thousand Oaks, CA: Sage.

Duffy, B. P. (1994). Use and abuse of internal evaluation. *New Directions for Program Evaluation, 64*, 25-32.

Fetterman, D. M. (1996). Empowerment evaluation: An introduction to theory and practice. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation* (pp. 3-48). Thousand Oaks, Ca: Sage.

Greene, J. G. (1987). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*(2), 91-116.

Guba. E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Thousand Oaks, CA: Sage.

Guba. E. G., & Lincoln, Y. S. (1981). *Effective evaluation*. San Francisco, CA: Jossey Bass.

Horton, M., & Freire, P. (1990). *We make the walk by walking.* Philadelphia, PA: Temple University Press.

House, E. R. (1978, March). Assumptions underlying evaluation models. *Educational Researcher*, 4-11.

Jaeger, R. M., O'Sullivan, R. G., Hecht, K. A., House, E. R., & Stake, R. E. (1986). Modular design for an evaluation of the Close Up Foundation programs. Sponsored by *RJR Nabisco*. Greensboro, NC: UNCG, School of Education, Center for Educational Research and Evaluation.

Jenness, M., & Barley, Z. A. (1995). Using cluster evaluation in the context of science education. *New Directions for Program Evaluation, 65*, 53-70.

Levin, H. M. (1996). Empowerment evaluation and accelerated schools. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment & accountability* (pp. 49-64). Thousand Oaks, CA: Sage.

Linney, J. A., & Wandersman, A. (1996). Empowering community groups with evaluation skills: The Prevention Plus III Model. In D. M. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation: Knowledge and tools for self-assessment & accountability* (pp. 259-276). Thousand Oaks, Ca: Sage.

Mansberger, N. B. (1993). *Making sense out of a mixed bag of data: The use of working hypothesis.* Paper presented at the annual meeting of the American Evaluation Association, Dallas, TX.

McNelis, R. H., & Bickel, W. E. (1996). Building formal knowledge bases: Understanding evaluation use in the foundation community. *Evaluation Practice, 17*(1), 19-42.

Mowbray, C. T. (1988). Getting the system to respond to evaluation findings. *New Directions for Program Evaluation, 39*, 47-58.

O'Sullivan, R. G. (1996). *Strategic evaluation design for the organizational development grants program.* Paper presented at the annual meeting of the American Evaluation Association, Atlanta, GA.

O'Sullivan, R. G. (1995). From judges to collaborators: Evaluators' role in science curriculum reform. *New Directions for Program Evaluation, 65*, 19-30.

O'Sullivan, R. G. (1984). *Evaluation in developing countries: A case study of the St. Kitts Teenage Family Life Education Program.* Unpublished doctoral dissertation, Auburn University, Alabama.

O'Sullivan, R. G., & D'Agostino, A. (1998). *How collaborative approaches promote evaluation with community-based programs for young children and their families.* Paper presented at the annual meeting of the American Evaluation Association, Chicago, IL.

O'Sullivan, R. G., D'Agostino, A., Prohm, B., Bundgaard, S., & Roche, L. (1998). *1997-98 Interim evaluation report of the Forsyth Early Childhood Partnership.* Greensboro, NC: Teaching to Diversity, University of North Carolina at Greensboro.

O'Sullivan, R. G., D'Agostino, A., Prohm, B., Roche, L., & Schmidt-Davis, H. (1997). *1997 Annual evaluation report of the Forsyth Early Childhood Partnership.* Greensboro, NC: Teaching to Diversity, University of North Carolina at Greensboro.

O'Sullivan, R. G., Clinton, J., Schmidt-Davis, H., & Wall, S. (1996). *First year smart start evaluation of the Forsyth Early Childhood Partnership.* Greensboro, NC: Teaching to Diversity, University of North Carolina at Greensboro.

O'Sullivan, R. G., & O'Sullivan, J. M., (1994). *Evaluation Voices: Promoting cluster evaluation from within programs.* Paper presented at the annual meeting of the Canadian Evaluation Society, Quebec, Canada.

O'Sullivan, R. G., & O'Sullivan J. M. (1998). Evaluation voices: Promoting evaluation from within programs through collaboration. *Evaluation and Program Planning, 21.1*, 21-29.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1988). Integrating evaluation into a program for increased utility and cost-effectiveness. *New Directions for Program Evaluation, 39*, 85-94.

Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelimsky & W. R. Shadish (Eds)., *Evaluation for the 21st century* (pp. 470-476). Thousand Oaks, CA: Sage.

Scriven, M. (1996a). Evaluation: The state of the art. *Supplement to the Proceedings of the Australian Evaluation Society 1996 Conference.*

Scriven, M. (1996b). Types of evaluation and types of evaluators. *Evaluation Practice, 17,* 151-161.

Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation, 58.*

Stevens, C. T., & Dial, M. (1994). What constitutes misuse? *New Directions for Program Evaluation, 64,* 3-14.

Stake, R. E. (1997). Advocacy in evaluation: A necessary evil. In E. Chelimsky & W. R. Shadish (Eds)., *Evaluation for the 21st century* (pp. 470-476). Thousand Oaks, CA: Sage.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R. E. (1991). Retrospective on *The Countenance of Educational Evaluation.* In M. W. McLaughlin & D. C. Phillips (Eds)., *Evaluation and education: At quartercentury* (pp. 67-88). Chicago, IL: National Society for the Study of Education.

Stake, R. E. (1983). Program evaluation, particularly responsive evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models* (pp. 287-310). Boston: Kluwer-Nihoff.

Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher, 7,* 5-9.

Usher, C. L. (1995). Improving evaluability through self-evaluation. *Evaluation Practice, 16*(1), 55-64.

Creating Evaluating Organizations

James R. Sanders
Western Michigan University

I wonder if Mrs. Hull (Sara's teacher) and Mr. Tykociner (the man next door) ever had the opportunity to participate in a program evaluation. There are many knowledgeable, experienced, talented people like them who are untapped natural resources in our communities. These are people who often "know" about programs in ways that the "experts" can never approximate.

How can we engage Mrs. Hull and Mr. Tykociner in our communities on evaluations of school programs, scouting activities, Boys and Girls Clubs, YMCA? How can we get them to think like evaluators--asking good questions, sharing information, using information to guide change.

We have a project in Kalamazoo, Michigan called The Greater Kalamazoo Evaluation Project (GKEP). This project was initiated by funders in this community--a private foundation, the community foundation, The United Way, and Community Mental Health--to encourage the use of evaluation in community agencies. Their intent was to communicate evaluation concepts in terms that everyone could understand and encourage community members to evaluate organizations and programs that are important to members of the community.

A task force of volunteers with an interest in evaluation was created and this task force provided guidance to a project staff hired to create:

1. An evaluation guide called *Evaluation for Learning*, which I am distributing to you.
2. Workshops that helped community agencies get started with evaluation.
3. Pilot projects in volunteer organizations that served to demonstrate ways in which evaluation could flourish.
4. Technical assistance for agencies seeking evaluation advice.

5. A newsletter that served to remind community members of the values underlying the use of evaluation. I am giving you two issues of the newsletter as an example of our work.
6. Discussion groups for those who want to talk about evaluation.

I want to share with you three cases of organizations that have found evaluation to be a positive energizing force in their development. The first is a community center with a staff of five that has used evaluation for internal planning, building staff morale, and for external marketing. The staff have kept evaluation simple, but true to values and principles of sound, participatory, open evaluation. One tangible benefit has been the incorporation of outcome thinking into everything they do. When someone wants to go to a workshop they can expect to be asked how it will relate to the organization's mission.

A local theater company is asking its audiences for feedback and is interviewing members of the theater community to check on the direction its board has planned to take. This is a small company of six board members and one staff member with an annual budget of $35,000.

Our local hospice director has said that she wouldn't do evaluation if it didn't pay off. This organization uses evaluation feedback from client families to guide improvements. Interdisciplinary staff teams are used to address difficult problems. They have found evaluation to be a morale builder. Make Us Great can be found on their coffee mugs.

The fact is, communities can become evaluating communities, beyond the usual commissioned or mandated studies. It takes a common mind set, community leadership, and perseverance. Mrs. Hull and Mr. Tykociner would be welcomed co-evaluators in Kalamazoo.

Helen Simons

Lou Rubin

Norm Stenzel

"Give Me An Insight": Training and Reporting in Naturalistic Evaluation[1]

Helen Simons[2],
University of Southampton

The title for this paper is a quotation from a colleague co-ordinating a development programme in Poland that is preparing Polish academics and teacher educators to evaluate educational transformation in their country. I am the external European Union consultant on the project. Our task is to establish an evaluation capacity in that country to the point where our Polish colleagues can evaluate independently without EU support. This context is particularly important as we shall see later, though the incident I am about to describe and the issues it raises affect us all.

We had been working intensively all weekend on "training" our "foreign" colleagues to observe. We had conducted several workshops which involved observing a mathematics lesson and observing a lesson in mastering a team activity utilising different forms of observation. These included a checklist, a criteria focused observation relevant to the task, narrative description, analysis of language and pedagogic analysis. This had been preceded by a previous workshop on listening and observing skills where similar exercises in watching and listening had been programmed.

The problem we encountered then, and in the experience I am currently relating, is that our "foreign" colleagues did not always observe what was happening. What they did was to offer "their" judgement on what was taking place, impute motivation to what actors were doing (with no evidence to substantiate the inferences they were making) and to describe in categories that were not relevant to the task--in effect to fail to "see" what was taking place. The point of repeating the exercise was to encourage the observers to produce relevant evidence-based observations.

---

[1] Revised August 1998

[2] This is a paper in progress. I would be pleased to have feedback on the issues presented here and would enjoy exchanging ideas for evaluation training. Please address correspondence to: Helen Simons, Research and Graduate School of Education, University of Southampton, Highfield, Southampton, S017 IBJ, UK
E-mail: hrs@soton.ac.uk Telephone: +44 (0)1703 593474

Giving judgement, rather than observing what is there, is not only an issue with training evaluators in Central and Eastern Europe. In evaluation training in many Western industrialised countries I have experienced the same problem. It takes a very long time for "novice" evaluators to learn to observe and to actually report what happened accurately, impartially and insightfully. To return to my story.

Each of the participants fed back their observations to the whole group--observers, participants and leaders of the workshop. We all listened and independently had the opportunity to check the validity of what was seen through the different methods by which the activity was observed and reported. It was better the second time around. More evidence was offered for the observations. Judgement was still inescapable for some. But there was some indication that the complexity of the task was recognised.

Reflecting upon the workshop later, the co-ordinator said of the evaluators:

> Some seem to be reporting more accurately but . . . I still feel . . . I am a little disappointed . . . they did not tell me something I did not already know. I mean, "give me an insight." That is what I am looking for.

This comment resonated with something Jackie Hill said to me years ago when conducting her case study for the Stake and Easley Case Studies in Science Education Project (Stake and Easley, 1978), "*I have to* interpret," she said, "*I cannot simply describe to them what they already know*."

The situation was not exactly parallel as Jackie was talking about consciously interpreting the data theoretically and signposting these interpretations for the reader, whereas in the context I have just described I am talking about unexamined, imputed and, often unwarranted, judgement.

My response in the Eastern European context was to say:

> I think I would be a little careful in asking for this directly at this point. You may not get what you are seeking. Given the state of the art of evaluation in these countries and the role of observation within it, what you may get is not insight but rather a judgement that is judgemental. I would stay close to the evidence for some time yet.

There are particular contextual reasons for making this response. In many countries in Central and Eastern Europe contemporary evaluation is a new concept. Pre-1989 any activity or outside interest in performance was largely equated with control and regulated output. Evidence from recent Central and Eastern European funded projects (see Hyatt and Simons, 1998) confirms that the dominant perspective of evaluation held by those with whom we worked was a particular characterisation of what we in the West call accountability (Chelimsky, 1995).

Partly because of this, there is, or was (the position is slowly changing with alternative experiences) a tendency to be suspicious or fearful of evaluation. This had two effects. One was to be suspicious of outside influences even though they were sought. The other was the avoidance of critique in the evaluations the participants conducted themselves. The fear of reprisal still held a force which the "foreign" evaluators managed in practice by not being critical of anything.

A third contextual point is the issue of judgement. Though fearful of other's judgements ironically, when taking on an evaluation role, some participants became very judgemental indeed--a case perhaps of reversal of power and roles. However I suspect that this had more to do with their authority as senior academics and the need to have their discipline-based expertise acknowledged and demonstrated to EU consultants and, in their "new" role as evaluators, to their peers.

"*Give me an insight*" in this context and how you train people to *"give it"* is quite problematic. This encounter led me to think about how in the context in question and in our own

---

[3] The particular characterisation of accountability that was dominant in these cultures was one associated with audit, exposure, criticism, inspection, legitimation. It did not encompass professional accountability or self-accountability.

contexts we prepare and "train" evaluators to "give insight." Can it be done? Can everyone do it? Are there contexts in which it is not advisable? Is it experience-related? What are the dangers in saying that it *is* experience-related? Is it something you can hand on or only encourage others to access or intuit? I will not have time to discuss all these questions in this paper but will focus on three: Is everyone capable of having insight? Is experience necessary? Can it be taught?

First though, what is insight and how do we recognise it? I am not sure definitions help. Meanings and contexts matter more. The common recourse to the dictionary (OED) yielded little help. Insight was characterised as "*penetration into character or circumstances with understanding*" (p. 519) and close to discernment. In indicating that the word discernment was *"insight, keen perception,"* the OED (p. 272) was slightly more helpful. Yet the second meaning given for discernment "*distinguish good from bad*" does not help. Insight it seems to me is something more directly perceived than this, more immediately grasped or felt and more holistic. The definition of intuition seemed closer still: "*immediate apprehension by the mind without reasoning, immediate apprehension by a sense; immediate insight.*" To intuit is "*to receive knowledge by direct perception*" (*OED*, p. 257).

On first sight this definition of intuition seems to be the same as the notion of insight I am exploring in this paper. Yet some participants at the seminar were keen to maintain a distinction between intuition and insight, reserving the former for the process of "intuitive knowing" that stems from previous cognitive reasoning and knowledge and the latter for the sudden recognition of something that makes sense of a complex situation, event or experience. It may be an intuitive, rational, or sensory process or a combination of all three.

Whatever definition one might choose, insight means different things to different people. It is one of those things, like quality, that we all know what we mean when we have it, recognise others have it or see it in their work. We all know what a "Stakian" insight is I expect. How can I describe it?--succinct yet complex, direct yet enigmatic, epigrammatic with layers of meaning. Master of the vignette, Bob has a characteristic way of conveying "insights" on the postcards he has sent over the years. Through the choice and composition of simple language, description and metaphor he presents a vivid portrayal--insight--in a few words. Take the following postcard for instance. It comes from Brazil:

> Dear Helen - Have enjoyed the resort, "radioactive" sand at this place, and just now have returned from a week visiting rural schools down the coast and into the mountains. Saw 21 1 room schools, Gr 1-4. Many pretty sad. Barren rooms, the dust and trash ever present. Teachers have workbooks for kids but no books, little paper. They carry water for the toilets, have no electricity, sometimes not even a woodstove to cook the pasta and beans govt. sends. County coordinator makes up final exam, sells it (15¢) to kids to cover office expenses, teachers buy when kids can't afford it. Kids have to get 80% right to pass to next grade, so some kids get more than 4 years of education. Yet spirits are high. Bob

What we are less sure about of course is how people come to have insight. Is it through their research, their life experience, something personal about them--or all of the above? Whichever, there are different routes one can take to gaining insight. I have identified four, though there may well be others.

First, there is the research insight we come to through a formal analysis of data, cross checked, referenced, validated and interpreted (through various theoretical lenses) to make meaning of events and circumstances described.

Secondly, there is the insight one gains through the direct perception that David Bohm talks about and, slightly differently, the "direct encounter" and challenge to conventional ways of seeing which Stake and Kerr draw attention to in their discussion of Magritte (Stake and Kerr, 1994).

Thirdly, there is the indirect, mystical insight, if you like, which the aborigines and other cultures draw on in following their "songlines" and related cultural traditions (see Chatwin, 1987). Some people refer to this kind of insight as "the silence within."

Fourthly there is the personal insight one gains from reading novels such as those by Virginia Woolf, *The Alchemist,* by Paul Coelho, or the New Zealand novel by Sonagh Koea, titled, *Sing to Me, Dreamer,* to name but a few examples. It is also experienced from engaging with the paintings of Cezanne, Magritte, Matisse to note some of my favourite painters; and the resonance we get from the images and insights in the poems, for example, of Raymond Carver, E E Cummings, Maya Angelou, Robert Frost; and from the short stories (vignettes?) of Katherine Mansfield. Biography is a rich source too, but less direct.

As researchers and evaluators we are perhaps more concerned with the first two ways above of reaching insight in our evaluations. However our research and reporting may well be enhanced if we were able to access more of the indirect insight (demonstrated in the songlines, for instance) in our own culture and utilize further what resonates through engagement with the arts.

Preliminary Answers

In this second section of the paper I will try and address the three questions on insight I raised earlier. I conclude with an attempt to devise evaluation training that will alert or awaken evaluators to different ways of gaining and revealing insight.

My answer to the first question **is everyone capable of having insight** is "yes." It has to be yes. I cannot subscribe to a view which claims some people have insight, or can come to gain it, while others inherently cannot. This is a different issue from saying whether it can be taught or facilitated which I will come to in a moment. The position I have to take, on egalitarian grounds, is that all people have the potential for gaining insight but not everyone develops the capacity or displays it.

Some choose not to use it for different reasons. It is not always acceptable to one's peer group to show insight and some may be fearful of possible reprisal if they do so. Take the "tall poppy" syndrome, for example, that has plagued

---

[4] The "tall poppy" is a phrase which I first came across in Australia. It refers to the phenomenon of the beautiful poppy being

many a schoolboy in the playground with corresponding repercussions on his school performance. The "tall poppy" syndrome has also been responsible for many women (and others) in certain cultures undervaluing their intuition and "feminine" ways of knowing.

Some do not believe that they have insight, a powerful inhibitor given that belief systems so often circumscribe (consciously or unconsciously) what we do.

Some do not know that they have insight, or the potential to have it because the education system frequently does not value the insights that happen through spontaneous interactions, unexpected encounters and apparently irrelevant observations. We are encouraged more and more to set goals, targets, outcomes: a) as though these can be attained and b) there is a route directly to them. Well there may be but, I suggest, at the expense of the "direct encounter" (Stake and Kerr, 1994), the "active participant" in observation (Rilke, 1991) and the acceptance of the totally unexpected. These are the situations and the contexts which allow people to have or to access "insight."

Is **experience, long term involvement in the field, necessary?** Here there is a two-fold answer. There is no doubt that in some contexts, especially those which are unfamiliar to us, deep or long term immersion in the field may be a necessary prerequisite to having accurate insights of that setting or people's actions within it. This is a point I was acutely aware of in working in Eastern and Central Europe both in my own response and in trying to "train" novice evaluators.

In cultures more familiar to us, insight may appear more readily, although the instant "insight" we recognise may also be an overworked metaphor, image or observation that has ceased to yield new understanding. In these situations we have to unlearn and/or learn to see in different ways. (See Stake and Kerr, 1994; Simons, 1996). So experience can have at least two dimensions--facilitating insight and, in some circumstances, preventing it. The important point for training

---

cut down by others when it grows (excels) too tall. Fear of this happening leads to under performance.

is the need to be conscious of how and when experience may be blocking insight.

In general I would say experience helps. Yet there are some people, irrespective of age or experience, who readily have insight, much as Gouldner (1973) would say there are some people who are simply more objective than others, and can be more objective about themselves. I think we can all recognise such people, many here today perhaps, who have this capacity for being insightful, in whatever context. As such they have a head start, I suggest in qualitative evaluation. Justifications, warrants and demonstrations will still be necessary. But direct perception from these people will be trusted more readily.

Towards an alternative training programme for evaluators

My third question is **can insight be taught?** Or, put differently, **how do we prepare novice evaluators to have insight?**

Training can take a myriad of structural forms from a six week course, a series of courses spread over two or three years, Masters and Doctoral modules, to a full scale apprenticeship in the field with an experienced evaluator. For some of us there has been a fifth approach. It's called *"being thrown in the deep end."*

Traditional evaluation training programmes of the course variety (at least ones I have been involved with) usually comprise an examination of different models, their epistemologies and promise, brief history of the evaluation field, issues of design and sampling, discussion of a range of methods, reading of seminal texts, different modes of analysis, styles of reporting, writing for different audiences, and ethics and politics.

There will also be attention to methodological issues such as validity, reliability and triangulation, time spent critiquing different examples of evaluation reports prepared for different purposes and, in some cases direct engagement with field data to analyse and present findings in different forms such as portrayal, case study, narratives, educational criticism and policy reports.

This form of preparation is all essential but I would want to emphasise two aspects of it that require more attention, than is currently given perhaps, in the search for "insight" raised in the quotation with which this paper began. The first is observation. Experience suggests that much more time is needed to help people to learn to listen and observe. Preparation should be multi-dimensional, involve peer critique, self-evaluation, triangulation of methods and by persons of the same event and public discussion of such observations. This is in no way to seek consensus but rather to see how observations are arrived at, how they may differ, and how they are justified.

The second is analysis. In all contexts I have worked in but especially in institutional self-evaluation and programme evaluation in Eastern Europe, much more time needs to be spent on different ways of analysing and reporting the data. Students and "novice" evaluators need to experiment with writing cameos, vignettes, portrayals, clear descriptive, accounts, highly interpretative accounts, narratives, and theory-led accounts to see what each of them communicate and how and whether these forms of reporting do encourage others to access the "insight" they may have gained in the evaluation. If naturalistic generalisation (Stake, 1978) is to work, readers of our evaluation reports need to recognise the scenario and context being described in a clear, vivid way to imagine and have the vicarious experience that will enable them to generalise.

My alternative curriculum for evaluation training would also include:

- an on-going built-in link with theory and practice e.g. the conduct of a case study, portrayal or policy report alongside formal "training" sessions utilising data from the person's own field work;

- time spent with experienced fieldworkers and evaluators working with rather than being *"apprenticed to;"*

- exposure to experiential ways of coming to know which would enable students to experiment with different forms and ways of understanding, take risks in creating alternative interpretations, to "dance with the data," and to have space to allow "insights to surface;"

- indepth course on the self in research and evaluation;

- readings, poetry, painting and music for the soul. I will not suggest a list here. You will all have your own but I would be delighted to exchange.

Conclusion

With such a programme we may get a little nearer to encouraging those who wish to evaluate in this way to "give or share insight" with others. In conclusion, however, I am drawn back to two points I raised earlier. Contexts matter and some are more receptive than others to accessing or revealing "insight." Take the evaluation for instance which Saville Kushner (1992) conducted of the Guild Hall School of Music in London, an institution which only takes first class, bright performing music students. On the front cover is a quotation from a student "*You are not just fighting the institution, you are fighting the dream*" which sums up in metaphor the experience of this student as well as telling a great deal about the institution. Much is due to the evaluator, of course, in creating the appropriate relationships which allowed such a perception to emerge and be voiced. But there is no doubt that the context of the institution and the articulateness of the student also had a strong role to play.

People matter. While I hold the view, as I said earlier, that all people can have insight, it is also true that some, through personal and/professional experience or simply because of who they are exhibit insight more than others. We have no better example here today than the person we are honouring. So I am content to leave the last insights with him. They come again from one of his favourite ways of communicating--the postcard.

Clarifying something I did not follow once came this reply "*I'm afraid I'm overambiguous.*"

Keeping me up to date, "*I was advised, with others yesterday in Washington, that some things should be left unsaid ... but he didn't say which!*"

Similarly, keeping me up to date *"I'm working on an anti-rationality paper for Evaluation Research Society meeting in Washington DC on Nov. 2nd, having trouble thinking rationally about it;"*

And finally (displayed postcard, glossy white on one side with semblance of postmark only). On the back was my address and the following message:

"Having a subtle
time. Wish
you were here.
Bob"

Well, we are here now--
Bob and we thank you for all the insights you have given us over the years.

Helen

References

Coelho, P. (1988). *The alchemist.* London: Harper Collins.

Chatwin, B. (1987). *Songlines.* London: Picador.

Chelimsky, E. (1997). Thoughts for a new evaluation society. *Evaluation,* 3(1), 97-109.

Gouldner, A. V. (1973). *For sociology: Renewal and critique in sociology today.* London: Allen Lane.

Hyatt, J. & Simons, H. (1998). Cultural codes: who holds the key? The concept and conduct of evaluation in Central and Eastern Europe. To be published in *Evaluation, The International Journal of Theory, Research and Practice.*

Koea, S. (1994). *Sing to me, dreamer.* Auckland: Random House New Zealand Ltd.

Krishnamurti & Bohm, D. (1991). Can insight be awakened in another? In *The Ending of Time, Thirteen Dialogues.* London: Victor Gollancz Ltd.

Kushner, S. (1992). *A musical education: Innovation in the Conservatoire, Geelong.* Victoria, Australia: Deakin University Printery.

Oxford English dictionary (The concise Oxford dictionary of current English) (7th ed.). (1982). Oxford: The Clarendon Press.

Rilke, R. M. (1991). *Letters on Cézanne.* London: Vintage Edition.

Simons, H. (1996). The paradox of case study. *Cambridge Journal of Education,* 26(2), 225-240.

Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher,* 7, 5-8.

Stake R. E. & Easley, J. (Eds.). (1979). *Case studies in science education.* Urbana: University of Illinois, CIRCE.

Stake, R. E. and Kerr, D. (1994, April). *Rene Magritte, constructivism, and the researcher as interpreter*. Paper presented to the annual meeting of the American Educational Research Association, New Orleans.

Possibilities For Cultivating Evaluative Intelligence

Lou Rubin
University of Illinois

Evaluation and intelligence are both abstractions but their connections are obvious. For example, astute judgement in appraising educational programs and processes, in order to uncover useful clues to improvement, is an indispensable element in enhancing learning outcomes. A major shortcoming of many evaluation training programs, however, is their tendency to convey principles, theories, and methods—ignoring the corollary skills essential to their successful application. In functional assessment—knowledge of method alone—is rarely sufficient.

Administrators frequently encounter difficulty in problem solving because they lack, first, the requisite knowledge of appropriate data collection and analysis; and second, what might be termed evaluative intelligence—the capacity for problem identification, interpretation, and resolution. If there is any validity to the notion of multiple forms of intelligence which can be coalesced as circumstances require—and if effective evaluation necessitates particular mixes of these forms—seemingly, it can be cultivated.

The need, certainly, is clear. The recent Rand study on school reform made it plain that the costly New American Schools Initiative has, to date, not brought much in the way of improvement and change--for a variety of reasons--but lackluster leadership, and the inability to distinguish the symptoms of problems from the problems themselves, rank high.

Potential correctives could readily be devised. Suppose, as a simple illustration, we organized a series of evaluative workshops designed to develop appraisal skills. At each session, a brief case study synopsis, depicting an educational problem, would be distributed and discussed. In the ensuing dialogue, the pros and cons of alternative evaluative strategies could be debated, particularly with respect to the essential information for a sensible analysis; the impact of contextual factors; and various pragmatic issues such as costs and benefits. Theoretically, the participants, over time, could develop broadened perspectives; collectively fashion operational rubrics; and, perhaps, even evolve assessment procedures which could be tried-out in situ; appraised; and revised. Admittedly primitive, the approach might, in one small way or another, enhance evaluative skills with respect to (a) deciphering instructional problems, (b) locating their source, (c) developing remediations, and (d) determining outcomes.

Accomplishing evaluation objectives presumably necessitates (a) obtaining essential information, (b) distinguishing which of it is of greatest significance, (c) organizing these critical factors into a usable matrix, and (d) balancing the resulting implications against pertinent insights derived from the evaluator's previous experience. Could we, then, not invent ways to hone and sharpen these capacities.

In somewhat the same connection, much has been made over the distinction between academic and practical intelligence (Sternberg, Wagner, et al.). Whereas academic intelligence involves the accumulation of knowledge through schooling--practical intelligence has to do with the adeptness through which tacit knowledge—intuitive understanding—is extracted from experience. If leadership intelligence can be nurtured through explicit exposure—there could be virtue in developing practical knowledge through specialized program provisions. The following basic assumptions may underscore illustrative points of focus:

- Evaluators continuously seek updated information to support their estimations.

- Evaluators--to enhance their understanding of phenomena--interpretively construct acquired information.

- Knowledge formulation—for purposes of evaluative intelligence--involves encoding, storing and retrieving.

- Since schooling is culturally and contextually bound, both must be considered in appraising outcomes.

- Evaluative intelligence is directly related to an action, its context, and the evaluator's professional sophistication.

- Information processing is a fundamental dimension of cognitive intelligence.

- Cultivated intelligence utilizes signs and indicators—that evaluators decode through a kind of semiotic constructivism.

- Since human activities are interrelated in a given situation, evaluation must consider the multiple aspects of cause and effect.

- Cultivated evaluative intelligence can be directed toward specific school improvements.

The real question, obviously, is what produces superior evaluation. The ongoing research offers some general hints: the best evaluators (a) make use of tested principles, (b) act upon judgment stemming from experience, and (c) use intuitive reasoning. They excel at analyzing consequences in order to make significant connections. Since educational phenomena are not always predictable, it often is necessary to alter strategies, try a different tack, or abandon one procedure in favor of another. It would be foolish, therefore, to expect evaluators to (a) follow prescribed steps, (b) function only in accord with established theory, or (c) adhere to predetermined plans. Moreover, the constructive use of hunch can be a useful tool. Good evaluators, for example, frequently rely upon shrewd discernment, gleaned from long experience in data-sifting, which has been processed and internalized through progressive reflection. For this reason, there is considerable danger in the spurious assumption that imitating expert evaluators produces expert evaluation. Imitation may enable apprentices to emulate useful procedure, but it does not guarantee excellence. The best evaluators are flexible in their approach: what they do, in sum, fits the circumstances at hand. Simply replicating standard procedure, without due regard for the reference frame, can heavily limit effectiveness. Thus, if a good tactic is used at the wrong time, or in the wrong way, the benefits are likely to be minimal.

It is not so much what expert evaluators do, but rather the ways in which they decide what to do that makes the difference—the logic with which they choose one tactic over another, or deploy conditioned instinct in choosing the best alternative. These decisions stem, in part, from looking for connections, interpretation, and extracting memory of past lessons which can be brought to bear on the problems at hand.

Knowledge and execution are independent. One can sense what needs to be done, for example, but lack capability. While academic preparation deals predominantly with knowledge, evaluation is an art that can only be acquired through exposure to the real world of practice and direct engagement. The rules of procedure, moreover, often must give way to the demands of a specific problem. Principles provide rules of thumbs which serve as guides, but finding the right rule of thumb is problematic. Hence, evaluation intelligence necessitates not only a consummate understanding of the schooling mileu but also prescience, and a portfolio of skills matured over time. The ability to recognize contextual constraints, ill-structured processes, and faulty information analyses are examples of such skills—acquired and honed through informed practice. The four principles which follow, derived from the literature, may help illuminate the place of evaluative intelligence in clarifying and utilizing generic guidelines:

> Formative evaluation, according to Scriven, is required when the objective is improvement. Accordingly, cultivating evaluative intelligence could be so oriented—since formative judgement is rooted in practice.
>
> For purposes of improvement, the use of evaluative intelligence should focus on casual factors, their conjunction, and the requirements for solution.
>
> Analytic evaluation--the auditing of select aspects or components--can be done separately, or in combination. Or as an alternative—global evaluation —a one-step, overall appraisal, can be employed. Logically, therefore, evaluative intelligence is involved in determining which approach is preferable in a particular instance.
>
> Stake, in the 1991 NSSE yearbook argued that:
> "Practitioners need to be told what to do." "An evaluator needs to tell us some things . . . convey some summary of findings—plus provide guidance in changing our

> practices . . . Not many authorities and practitioners may be persuaded—or even take heed—but the responsibility to give counsel exists."

What this means, self-evidently, is that evaluation is a form of problem solving and its greatest good, therefore, lies in generating pragmatic improvements. Thus, it is the interpretation of collected evidence that is the marrow of evaluative intelligence, and most compelling. The evaluator's mission is not merely applying formulas, but rather generating understanding which points the way. It is for this reason that the construal of events, contexts and intents are of great essence. Algorithms have their place, but, through means-ends analysis, it is the inspired intuition, in the form of a heuristic, that is likely to bring small quantum leaps.

Through their evaluative intelligence, evaluators should help us by clarifying what was right or wrong; suggesting better possibilities; monitoring progress; and reminding us when reconsiderations are in order.

Evaluation is not Evaluation is not Evaluation

Norman Stenzel
University of Illinois

Evaluators these days are spending considerable effort to professionalize their endeavors. Those labors include the development and publication of standards and considerable discussion about the notion of standardization of training and credentialing. In many ways such work presumes underpinnings that are not yet clearly defined. And, to the extent that clarity fails, the standards, delineation of preferred training and credentialing are potentially flawed or at least unsettling.

It is in respect to standards that the tract included here, Evaluation Is Not Evaluation Is Not Evaluation, authored by a Coalition (no date), presumes to reflect ideas derived from Robert Stake. While it may not be the case that the tract is an accurate reflection of Stakian teaching, the contention that there are quite different valuations reflected in quite different practices that are considered to be evaluation and that there are consequences of those differences has long been a concern of the denizens of CIRCE.

I will reflect upon the relationship between the nature of valuation in evaluation and the implications of multiple valuations for the *Joint Committee's The Program Evaluation Standards* (1994) in the following paragraphs.

Valuation and Evaluation

At one time, in the 70s and 80s it was common for authors to devise tables to characterize types of evaluation. (While I will not formally cite references here, I believe I recall one from Egon Guba, Ernie House, and even one from Robert Stake.) Such tables often were presented to provide comparisons of different features of the models under review. As I looked at those materials at that time I often wondered why some of the items were included, for it seemed to me that there was considerable redundancy and that the nature of the "types" were not clearly represented. It seemed to me that the differences were based on form and not substance. Now, the Coalition in their tract has made a distinction that makes sense to me. The Coalition lists four different conceptualizations of valuation embedded in what are recognized as common evaluation models.

Evaluation. This type of evaluation is exemplified in Stake's *Countenance of Evaluation* (1967). Value is established by the instrumental or contributive relationship between a transaction and what was intended to be the product of the action. That is, the doing of something leads to or, along with other transactions, contributes to an anticipated outcome. What was done is good if it works as was expected. Efficacy is a good making quality.

Evaluation2. Evaluation2 in the Coalition tract is a fictitious evaluation model patterned after a discrepancy approach. Evaluation2 is a variety of evaluation that in the tables of old, looked very much like the Countenance model—Provus' discrepancy model (1969). The Coalition, however, points out that the valuation provided through the discrepancy model is an assurance that one is getting the work that was contracted. It always seemed to me that this was more like monitoring/auditing than evaluation and that the valuation in such cases was provided in the proposal review by reviewers for funding. Yet now, I will allow that getting what one pays for is a value of worth. Not like getting your money's worth or true value for an investment--Evaluation5--but a value just the same. Keeping a promise/fulfilling a contract/living up to one's word is a good making quality.

Evaluation3. This type of valuation is found in accreditation models. It includes the considered discriminations of professionals of repute to identify the merits of an evaluatum. While the Coalition focused on accreditation models, the writings of Elliot Eisner about connoisseurship suggested greater breadth to this approach to valuation. Indeed, even Louis Rubin at the Stake Symposium in presenting a call for evaluative wisdom seems to support Evaluation3. Passing inspection based on sage experience has a long tradition in education and is a good making quality.

Evaluation4. The final valuation type identified in the tract is that depending on statistical differences. The tract indicates that value is established by measurable differences in outcomes--differences that are statistically significant. Evaluations in this mode have been able to use increasingly powerful tools, moving from simple differences between mean scores to complex studies including many variables made accessible through computer technology. Statistical difference is a good making quality.

Evaluation . . . While the tract itself does not provide explanatory narrative for the ellipses in the heading, we expect that the authors of the tract intended to indicate that their list could continue on. The implication is that there are other varieties of good making qualities. Whether or not we should agree with the tract that since there are quite different bases for valuation, a common set of standards is not possible, is yet another matter.

Valuation and Standards

One reading of the Coalition tract might be that there should be a variety of standards more specific to such different evaluation approaches as were enumerated. After all, that is often the basis for critical review of work in the research professions. If we are to determine the adequacy of an evaluation based on compliance with a contractual agreement or one more experimentalist in construct, there may be standards that should be added to the generic compilation included in the Standards. I then would have to agree with the Coalition that if more specialized standards are used to determine the strengths or weaknesses of an evaluation, the generalizability of a standard would be limited.

The fear that I would have about such a practice is that evaluations could get mired in an infinite cycle of challenge and response. Indeed, in the *Standards* the call for metaevaluation could presage more and more doubt about the credibility of evaluations rather than security derived from review.

So, do the *Standards* help us in considering the valuation in evaluation? *The Standards* does include a section on Values Identification under Utility Standards (pp. 44ff). That section urges evaluators to consider alternative interpretative bases, to consider who will make interpretations; to consider alternative techniques and to report options. Among the items in the "Common Errors" section that follows, there is some elaboration on the possible types of evaluation perspectives--educational, social, economic and scientific.

This section seems to form the basis of a counter example to the Coalition's claim that values are not well attended to in the *Standards*. Yet the Coalition may reply that it is not that valuation is absent, but that there are problems in determining how to deal with application of standards to the great variety of value perspectives possible in evaluative enterprise. Let me take a cue from the *Standards* and use an "Illustrative Case" to explore this matter further.

Illustrative Case—Description

Country School has been running a program for potentially truant students for a number of years, and local as well as state officials decide that it is time to determine the worth of the program. Local officials want to have an evaluation that will be suitable to inform the school board, parents and a number of advocacy groups. State officials want to determine the efficacy of the project with an eye to consider replication with special funds from the legislature, where members are considering school improvement funding.

The local officials hire E. Gunn to conduct a responsive evaluation. E. Gunn brings his teenage son along to participate in the examination of the project. They meet periodically with the school board representatives, local groups involved with truancy issues and parents of students in the program to obtain impressions of the interests that will need to be served in an evaluation report. The Gunn group conduct a number of observational activities to become acquainted with the implementation of the project. The elder Gunn follows that with interviews of administrators, teachers, and parents; as well as review of extant progress data. The younger Gunn sits in on classes for a week, reads materials, does assignments, talks with students, and interviews counsellors. The Gunn Reports (several for the several audiences) have been first reviewed by staff and students with comments having influence in revisions or attached as explanatory notes. The Gunn Reports conclude that while the teachers are industrious and results appeared respectable, the materials are mind numbing. In fact, students want very much to get out of the experience and back to their regular environments. A majority of students, however, indicated that they still would drop out of school as soon as they reached school leaving age.

State officials sent a team of reviewers to interview administrators, school board members, teachers and parents. They also sought out local judges, prosecutors and truancy officers who had made use of the program. Comparable interviews were conducted using prepared schedules. A portion of the team reviewed local data, collected data via the state academic assessment instrument currently being administered to students statewide. The State report indicated that the program was a tough minded approach to at risk students that had remarkable statistically significant success in improving the basic academic skills necessary for functioning citizens. The report provided to local authorities commended them for their enterprise and suggested that they apply for a dissemination grant. State officials were alerted to look for this promising replication project.

Local officials are amazed but are told to look to the *Standards* for guidance by a local expert in evaluation.

Illustrative Case—Comment

While the *Standards* provide a variety of admonitions in such instances as Utility Standard number 4—describe perspectives, procedures, and rationale used in interpreting findings--nothing prepares evaluators or consumers of evaluations to deal with the contrasting valuation illustrated here.

It might be that another CIRCE alum, Bob Wolf, would have proposed to use a judicial model to allow the confrontation of the disparate results in a setting juried by stakeholders. Wolf's approach certainly could require the Justified Conclusions called for in the *Standards* (A 10). Such a setting might have brought the student perspective to the attention of the State team and called for their response. They could have responded that students who are succeeding in school have a greater possibility of completing their education, and that boring content and statements of intent are countered by the improved academics.

There is another Utility Standards that could be of interest--information scope and selection (Utility 3). The State might be faulted for focusing too narrowly on information for replication. Yet they might have even proceeded with replication efforts if they justified their reliance on academic outcomes over sympathy for learners as illustrated in the Adversary Evaluation suggestion.

But, as I see what the Coalition was driving at, the document fails to guide us where valuation from a variety of evaluative perspectives are in conflict. Where in the *Standards* is there advice about, or even warning about conflicted judgments? The *Standards* attend to other topics in considerable detail while valuation is in the singular. The Country School project is and is not a success.

It would be easy to apply a qualitative standard to parts of the Gunn evaluation and utilize a Quantitative standard more heavily in the State evaluation. Yet we have little guidance in the Qualitative admonitions about the individual as an instrument. The use of the experiences of the younger Gunn are better understood and better judged from the perspective of "heuristic research" as described by Clark Moustakas (1967). In the State perspective, the focus on academic outcomes while discounting student perspectives might be judged by fully informed stakeholders as in an adversary hearing. These issues regarding valuation are beyond the existing standards, but vital to the future of our profession.

Back to Valuation

It seems to me that the existing standards count on method and procedures to be the basis for the valuational claim, "I have done this in this way and therefore I can make this interpretation." This is not sufficient. We need to be able to provide guidance about value claims based on disciplined inquiry of various sorts--perhaps we can look to House's *Evaluation as Argument* as a starting point. We need to be able to provide guidance when value claims from different sources appear to be in contention--in addition to the possibility of an adversarial proceeding. We are not much beyond the "You may say that, but I say this " stage of argument. Perhaps we will have to more thoroughly examine justification such as in the work of Carl Wellman (On Justification, 1988). We will need to consider valuation much more thoroughly to become the profession of our aspirations.

ATTACHMENT A

EVALUATION1 IS NOT EVALUATION2 IS NOT EVALUATION3 . . . 6

What We Learned From The Stake
Almost Agreed Upon by
The Coalition of 2 Plus or Minus 1

It is not part of the game in academe to admit that one has ripped off ideas from a venerable. Yet we see around us, especially in evaluation, the work of copyists. Think of the halcyon days of the late 60s, 70s and 80s. Those were days when it was common to borrow ideas from The Stake, give them a twist or a new label and call them one's own. This is our confession. We admit it. We had few, if any, original ideas. We ripped off The Stake big time. (Not that it did us any good. The Stake became famous. At best, we became infamous.)

Take *The Countenance of Evaluation* for example. We did. The Stake suggested we look for congruences between intentions and actual transactions and then determine contingencies of performance/non-performance-- Evaluation1. We ripped off that format in our OOOOPS (Operationalization Opportunities Of Objectives Purposefully Scrutinized) evaluation model--Evaluation2.

We advocated auditing governmental programs against their proposals. (It may have been that Provus borrowed his discrepancy evaluation model from us. Those are the breaks for rippers such as we.) We did not get a lot of jobs with our model. We were not sure why. The Stake prospered.

We had another encounter with ideas from The Stake. He was headliner at an inservice for the inner sanctum of the North Central Accreditation Association. He seemed to be quite supportive of the professional judgment version of evaluation—Evaluation3. We, then and there, decided to jump on board the professional review strategy of evaluation. So we created our own version of the accreditation style to apply to a variety of institutional type settings. While we were doing that, The Stake was moving on to portrayals and stakeholder issues. There we were, supporting the in crowd and The Stake was providing empowerments to the masses. We did not get a lot of work.

Next, we had heard that The Stake had been a trained quantitativist. We did not take time to verify that dark secret about The Stake's background, but made our move into the comparative statistical significance mode of evaluation—Evaluation4. We were bound to beat The Stake to the bounty in one area of evaluation. The Stake did not follow. We found that field crowded, and we did not prosper.

So there is our confession. Envy, greed, and arrogance led us to be the leading Stake rip offers in the nation.

Yet it has not all gone for naught. We have learned from our experience. These versions of evaluation--Evaluation1, 3--are not all of the same cloth. The nature of the valuation embedded in each process is quite different: Evaluation1 in The Stake's version is based on instrumental or at least contributive value. The OOOOPS version, Evaluation2, indicates that there is value in fulfilling a contractual agreement--the funding agency gets what it paid for. And, as we have pointed out, the judgment of professionals based on their best ken, Evaluation3, is the essence of accreditation models. The quantitativist quest for significant difference, Evaluation4, is well known. These evaluations are clearly not the same. They will not serve the same purpose, they will not lead to the same positive and negative valuations, and they cannot be judged by the same standards. (We call this our learning, but it may again be another rip off from The Stake.)

Readers everywhere, join the Coalition of 2 Plus or Minus 1 and confess your rip offs from The Stake. It will do your soul good.

References

House, E. (1972). *Evaluation as argument*. Los Angeles, CA: UCLA.

Joint Committee. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Moustakas, C. (1967). Heuristic research. In J. Bugental (Ed.), *Challenges of humanistic psychology* (pp. 101-108). New York: McGraw-Hill.

Stake, R. (1972). The countenance of educational evaluation. In Taylor and Cowley (Eds.), *Readings in curriculum evaluation* (pp. 93-102). Dubuque, Iowa: Brown.

Wellman, C. (1971). *Challenge and response: Justification in ethics*. Carbondale, IL: Southern Illinois Press.

Richard Jaeger

Tom Fox

Jacquetta Hill

Katherine Ryan & John Ory

Setting Performance Standards for National Board Assessments: A Reprise on Research and Development

Richard M. Jaeger[1]
University of North Carolina at Greensboro

When the National Board for Professional Teaching Standards began its teacher certification program, existing methods for determining appropriate standards of performance (e.g., Angoff, 1971, Ebel, 1972, Jaeger, 1982, Nedelsky, 1954) could not be applied to the Board's assessments. Most of the standard-setting methods in popular apply solely to tests composed of traditionally-scored, selected-response items. Indeed, the method due to Nedelsky (1954) can only be used with tests composed of multiple-choice items. These methods are inapplicable to the kinds of performance standards used by the National Board for Professional Teaching Standards for several reasons: they assume unidimensional, summative scoring of tests; they apply solely to dichotomously-scored test items; implicitly, they rely on the unbiasedness property of the Central Limit Theorem to average the judgment errors associated with individual test items. Once again, new measurement methodology had to be developed.

Beginning in 1991, the National Board for Professional Teaching Standards sponsored an intensive program of research on the development of standard-setting methods that are appropriate to its complex performance assessments. The research is still ongoing. The progress achieved through the National Board's research on standard setting has been reported regularly at meetings of the American Educational Research Association and the National Council on Measurement in Education and through the journal literature (Jaeger, 1994; Jaeger, Hambleton & Plake, 1995; Jaeger, 1995a; Jaeger, 1995b; Plake, Hambleton & Jaeger, 1995; Plake, Hambleton & Jaeger, 1997; Putnam, Pence & Jaeger, 1995).

---

[1] Editor's Note: Dick Jaeger acceded to Bob's request that he speak twice on the first moring. He spoke casually conversing with each group. This more formal presentation was taken from a paper Jaeger was developing at the time, "Setting performance standards for National Board assessments: A reprise on research and development." It was scheduled to be included in a volume of the JAI series, *Advances in Program Evaluation*, edited by Lawrence Ingvarson. He had worked on it while a Fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford University. He asked us to express his gratitude for financial support provided by The Spencer Foundation under Grant Number 199400132.

Three alternative standard-setting procedures have been used with the National Board's assessments since 1991. The Dominant Profile Judgment Method, described in Plake, Hambleton & Jaeger (1997), was originally developed by Jaeger and later refined by Hambleton and Plake. The method was applied only to the National Board's Early Adolescence English/Language Arts assessment, one of the initial two assessments developed by the Board. It required panels of standard-setting judges to specify the lowest profile of performance on the exercises that compose a National Board assessment that should result in candidates receiving National Board Certification. All candidates with profiles of performance that dominated the specified minimum (in the sense of having score values equal to or greater than the minimum) also would be certified.

The Dominant Profile Judgment Method resulted in the specification of a complex, multi-component performance standard. For example, to be certified a candidate would have to achieve a given total score across all exercises in an assessment, *and* achieve at least a specified minimum score on a subset of exercises considered by panelists to be most critical, *and* achieve a score greater than one on each of the exercises in the assessment. Although many standard-setting panelists appreciated the flexibility afforded by the Dominant Profile Judgment Method, this approach to standard setting was abandoned when it became clear that the complex performance standards it produced substantially reduced measurement reliability and, in particular, dramatically increased the probability that false-negative errors of candidate classification would occur.

The principal weakness of the performance standards produced by the Dominant Profile Judgment Method was their partially conjunctive nature. Whenever certification of candidates depends in part on their performance on a single assessment exercise, as with a standard that prohibits earning a score below some threshold on any given exercise, resulting reliability will be low. Regardless of the method used to derive them, conjunctive standard-setting rules--that is, rules that invoke multiple hurdles to achieve certification--should be avoided for this reason.

The standard-setting procedure applied most frequently to the National Board's assessments was termed Judgmental Policy Capturing. The method is described in a number of papers by Jaeger (Jaeger, 1994; Jaeger, Hambleton & Plake, 1995; Jaeger, 1995a; Jaeger, 1995b), its principal architect. When the National Board's assessments were expanded to include ten assessment exercises, the Judgmental Policy Capturing procedure had to be modified so as to present a judgment task that imposed reasonable cognitive demands on standard-setting panelist. A two-stage procedure was used for this purpose.

*Editor's Note: Jaeger passed out sheets, see pages 176, 177, for a judgment processing exercise through which he guided the audience and described the two-stage Judgmental Policy Capturing procedure.

Standard-setting procedures are, at base, methods for eliciting the reasoned judgments of qualified experts on test scores or levels of assessment performance that warrant some valued classification of examinees. Performance-standard-setting methods vary in the size and composition of the panels of experts used, in the training of panelists and the stimuli used to elicit panelists' judgments, in the decision aids used to inform panelists' judgments, and in the procedures used to compute performance standards from the judgments elicited.

The NAEP Definition of BASIC Proficiency in Mathematics

Fourth-grade students performing at the BASIC level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content strands: (1) estimate and use basic facts to perform simple computations with whole numbers, (2) show some understanding of fractions and decimals, (3) solve simple real-world problems in all content areas, (4) use--although not always accurately--four-function calculators, rulers, geometric shapes; (5) written responses are often minimal and without supporting information.

Problem 1

N stands for the number of stamps John had. He gave 12 stamps to his sister. Which expression tells how many stamps John has now?

A. N + 12
B. N - 12
C. 12 - N
D. 12 x N

D. 12 x N

Problem 2

Ms. Hernandez formed teams of 8 students each from the 34 students in her class. She formed as many teams as possible and the students left over were substitutes. How many students were substitutes?

Answer: ____________________

Problem 3

Sam can purchase his lunch at school. Each day he wants to have juice that costs 50 cents, a sandwich that costs 90 cents, and fruit that costs 35 cents. His mother has only $1.00 bills. What is the least number of $1.00 bills that his mother should give him so that he will have enough money to buy lunch for five days?

Your Calculations:

Problem	Probability
1	
2	
3	
Total	

The Legacy of Centers

Tom Fox
National-Louis University

Introduction

When it comes to legacies, about all I know is that gray haired people talk about them at times like this. And what is a time like this? A group of people getting together to reflect on a person's work over his (in this case) life, on influences, intended and unintended, on contradictions, intended and unintended, on the spirit of the work and the ways in which that work, and person, have been reinterpreted over time and in other lives. Another feature of this time is that the person is still among us, and so speaking of legacies is, perhaps, a will to bring the future into the present for our collective celebration of what can be accomplished in our limited, finite times. Being the timid, gray haired soul I am, I decided to consider the legacies of a professional fiction that many of us have lived, including Bob, only somehow Bob has carried this fiction off far beyond anyone's wildest dreams, including perhaps his own. That fiction, of course, is the notion of "centers," in this particular case, the Center for Instructional Research and Curriculum Evaluation, or CIRCE, the imaginary home of Bob Stake, and others over many years. Now, I don't know much fact about this center, except what I learned from its letter head, first seen when I met Bob through a CIRCE sponsored workshop/conference in 1976. Before that, I started a center of my own in 1974, and after that, I worked at the Centre for Applied Research in Education, CARE, for a few years, and learned among other things, of course, that "r" comes before "e" in the centre of real English. There have been, I am sure, thousands of centers which have served educational interests, so CARE and CIRCE are just two of a huge mass of centers begun in the last, say, since I entered education 39 years ago.

I don't have the attention span to review the recent history of centers, or even of CIRCE, unfortunately. Maybe you don't either. Instead, I take this opportunity to express a sort of eulogy on centers recognizing, of course, that centers are dead. (Can you imagine anything be labeled the center for action-research, for example, or the center of applied research in education?) I want to consider the legacy of centers because their variety, ambition, liveliness, intentions, and contexts were varied and rich for the brief geological instant of their existence at the end of modern times. A center of anything in our post-modern, deconstructive, post-structuralist 21st century future is a no-no. There are no centers of anything, anymore. (Although we may still sprout "centres for deconstruction," my guess is the irony isn't lost on anyone.) So this may be a reasonable time to consider what it is we can learn from this simply named phenomenon, and maybe entertain alternatives that capture professional needs in our centerless futures.

Let me return to Bob for a minute, since his person may not be irrelevant as we consider the notion of centers. Nor would the horrendous variety of persons who started centers in addition to Bob, for example Lawrence Stenhouse of CARE. In Bob we have a multidimensional personality, contradictory, consistent, edgy, interconnected, enigmatic, selectively iconoclastic, singular, maybe even lonely, socially-minded, a frontier-pusher, a traditionalist, a "saver" of values, a destroyer of former principles of action, and a creator of new ones. I am sure you all can add (and subtract) a variety of Bob's characteristics that go well beyond my own understandings and experience, and extend the singular complexities of the person. I think I could say something similar about the complexities of Lawrence Stenhouse, using, of course, quite a different set of contradictory characteristics. "Centering," in other words, may have been partially a personal need as well as a professional requirement for the unbounded folk who created and sustained centers. Acknowledging personal complexities stimulates my wish to locate some of the energies that have emanated from the institutional deceits of "centers," the professional bull's-eyes set up for some of the Robin Hoods of education to aim their ambition. The biggest deceit here, however, will be my attempt to generalize the legacies of centers, to try to raise some essentials passed from our center-age recent past. I would like to share the responsibility for the performance of this deceit, so anything you can add, or subtract, by interrupting me would be much appreciated.

Centers

First, centers were a place, an oasis, or at least a piazza, a plaza, a square in the midst of an otherwise ongoing professional community of discourse(s). A certain notion of space was tightly implied by the word "center," as fictional as the space often was. A robust idea needed a place, however finite, to stay, to grow and be maintained, to develop and be sustained by work. Space, of course, had a particular meaning over these years during the cold war, a place of identity and style rather than power, like Berlin. One didn't need a lot to have what was needed. The power of space had changed since medieval times, or since the two world wars. I saw a recent analysis by an Icelandic art critic (Olafur Gislason, 1998) of the centerless designs of many Icelandic towns today. He compared these no center towns, designed over the last few hundred years with their notion of disbursed, subsistence farming, to the Mediterranean and classic design of city-states with center squares. Many Icelandic towns have no center, which was a natural and spatial conception in their rural society. One conclusion he reached, as an art critic, was that these towns with no center had no need for fine arts. Instead they stayed with the narrative, and similar ties with nature. (When reading this analysis, I thought of the impact of rural America on American research in the 20th century, with many creators of centers having grown up on midwestern farms, with their own experience of centerless space.) Our city centers have had their own battles through modern times, with the steam engine, and then the car breaking up old town centers, creating alienation at the edges. Now, there is the reduction of other abstract borders as well, breaking down boundaries between private and public, for example. Gislason refers to the post-modern hyperspaces of glassed in structures, palm trees growing in northern climates, spaces that "transcend our capacities to locate ourselves," inventive structures that provide enchanted simulations, releasing us from the real, rejecting the central perspective of the renaissance in pictorial terms, or of the classical in political terms. There is a building like this in Reykjavik, now, where the notion of rural centerless space is coming in through post-modern architecture. Post-modern art, on the other hand, Gislason suggests, is fighting the multinational centers of capitalism by becoming unsaleable, independent of consumption. A certain kind of consumption, of course, was expected by centers, in fact, they were designed for public consumption in a way that other academic structures were not.

Centers were often set up to develop alternative styles of investigation, or to link inquiry with knowledge, research to curriculum, not only for the profession but for the clients, for those paying for the services. You knew if you went to a center, whether as student or another short-time visitor, you would enter a different sort of professional community. This was obvious first by its outward look beyond the academy, as well as by its own peculiar curiosity, its own doubts and certainties, committed together with common concerns (with or without a manifesto), supportive to one another's work, rubbing off sometimes on those who visited to be infested elsewhere. Intellectual vitality was often sustained within a maverick mode, accompanied by a frontier type bravado and joy. Centers, then, were places to go to, to identify with, to be infected by, to leave. Within a short time, they were considered as a constant, as a feature on the professional landscape that you knew would be there for guidance, like a beacon, even if other waters were being sailed. The center as a professional constant was a myth of course, better known by those within than those without, since most centers were, almost by definition, and certainly by funding, temporary, supported primarily by short-term grants and contracts. This made them more political in a non-ideological sense, linked to the public economics of the times, and tied often to specific or at least a small number of persons who could open the few money bags available for their survival.

So centers were always in a battle between permanency, the employment pattern of higher education, and the vagaries of public/private funding. Their necessary styles of entrepreneurship, in fact, brought them into realms of collegial disrespect and jealousy that made them nearly heroic to themselves, as well as to a few others, especially students. How centers have dealt with these professional, institutional and economic tensions should be a rich feature of their legacy. The specific strategies applied to gain their respective identifies, linkages, studentships, research and evaluation projects, especially evaluation projects, could tell postmodernists much about how to roam fields of decentered professional investment. It is interesting to me, for example, that the center movement was primarily in the fields of evaluation rather than curriculum development. Might that change?

A corresponding fact is that centers were never formed to be in the center of anything. They were formed instead to link the frontiers, the fringes, the boundaries of research experiences to those not initiated into research. Centers were formed to exist at the borders of education, not at the centers. They were expected to be off-center, eccentric, examples without originals--"simulacra," as Stronach and McClure, (1997) titled them, referring to the nature of the post-modern. Centers were often formed to push the boundaries of an expert field closer to those far away from the centers of a domain. In this way, especially in evaluation, they seem almost post-modern, deconstructivist organizations. Yet centers were certainly as modern a feature on the academic landscape as institutes, those collections of like-minded, expert experts, formed to talk to themselves, to sustain the esoteric depth at the heart of their own domains. Unlike institutes, most centers I knew were formed to push the edges of a field and simultaneously deliver those edges to the uninitiated. This may be one reason educational evaluation became populated with so many centers, and curriculum did not. Even as centers tried to make their inquiries accessible, it is interesting to me that Stronach and McClure's challenge for post-modern research could well have been hidden inside the desk drawers of center staff. "Let's see how far [educational research] can get by failing to deliver simple truths."

Centers carried a tendency to protect, a common fixture of any space claimed at a border. Some centers conducted their work as forts in the wilderness. The variety of ramparts, the thick walls of rhetorical protection are part of their legacy. Regardless, much of what centers did made sense in terms of their survival. They also made sense in terms of the times and circumstances, in a professional world that broadcast better from tightly constructed locations.

The Death of Centers

With the above virtues, and many more I may not understand, why should centers be dead now, especially when their realities, rather than their deceits, are closer to post-modern sensibilities, desires, and (even more powerful) rhetoric?

Perhaps centers were too successful as educational entities, as well as inevitable failures. The following are some conflicting outcomes, mostly unexpected and unintended. First, most centers showed very early on just how quickly the "untrained" could be highly productive. Thus, unlike academia, experts in centers weren't really needed by those who worked with them and left. Second, the entrepreneurial successes of centers co-opted the ambitions of the young who stayed. The excitement of the work made it difficult for the initially ambitious to leave, or even to expand beyond the external, impossible challenges of the next contract. Third, the greater the outreach of the centers, the more the competition for monies, the more possible were alternatives to their particular place on the edge. Outreach, then, could best be achieved by rumor with less unfairly naive competition than through normal channels of professional news (read mostly by the naive). Fourth, the resiliency with which centers continued to be consistent to their initial missions meant that offspring who flew from the nest had only one place to perch. Similar work performed away from the center was seldom considered as pure. There remained only one center. Fifth, retaining the educational ambition to create alternatives, the ambition to foster, stimulate, support, mentor, nurture, real alternatives to themselves, made the centers failures in their own standards. As educational seeds, centers recognized they had no fertility. Two external realities made this nearly inevitable (to say nothing of the internal ones already mentioned). The tenacity of those at the real centers of academia to retain and protect their own secure and narrow power made it certain that few alternative centers could be maintained, supported, or recognized in higher education. Furthermore, the inadequate resources, the slim economics for public works in the 80s and 90s made it impossible to support new centers, or even many of the old ones. Some may claim it is only the later that closed the windows of opportunity for the growth of new centers, but I would guess that a closer look would show how the friends of those who formed the centers made their eventual demise inevitable, regardless of the economics.

The Legacy of Centers

So what is the legacy? Are there any lessons that may be drawn about how to work at the edges in our post-centered times? Perhaps the first lesson is to realize just how well those who created the centers understood the modern era. They knew how centers broadcast out, they knew how singular identities could help retain their fame, how claiming a space in the middle was perceived as strength. They were, in other words, a modern fixture, just as significant to the modern sense of professional identity and performance as railroad stations were to the corresponding identity of towns and cities. This legacy would suggest a new sense of space may help, a response to a center-less space at the amorphous boundaries of multiple domains. Perhaps locating at a university (or multiversity) is more foolish than setting up in schools, for example. And setting up in schools may be more foolish than setting up in malls, or in churches. Or setting up anyplace may be more foolish than setting up chat rooms, or web pages, which have their own conceptions of connecting in non-space.

If we don't need space, what do we need to communicate in the post-modern world? Probably time more than anything else. Perhaps we can look at time, rather than space, for a new metaphor, our next deceit for focusing and distributing outward our energy, intelligence, visions. When it comes to time, of course, the university has taken over employment patterns from the centers by hiring adjuncts (instead of research associates) for a year (or less) at a time, with no institutional affiliation, support, or responsibility for welfare. Centers had belief and mission to hold onto their workers for the short-life of a contract, but the image, now, of the professional homeless (an image not far from many centers) is too close for comfort. In the center days, at least, money bought professional time, finite as it was. Today, one has either money or time, never both. Money buys no time for a robust idea, and those who may have time for developing a new idea into action, have no money. What is an idea person, an iconoclast, a frontier pusher, to do if she wishes to extend the edges, to crash the barriers of current orthodoxy's (many created by the centers), to push past the boundaries as she finds them in the 21st century? She can't form a center, so what is there? She might consider the flower rather than the greenhouse. Or time rather than space. She may try to organize an edge-like organization, something at the fringes more available and fluid for professional channel surfing, dipping in and out of existence, more mobile than centers ever were, an creation of multiple identities rather than a single identity. Or she may go to other features of the legacies of centers and see where they may take her. Let's try to identify some of those.

One legacy is the persons who founded centers, how they read their times, responded to them, and created certain physical attributes to carry their professional dreams. She may go to the people who formed the centers, or to those who have continued to maintain them through the recent rocky fifteen years, to hear what it took in personal sacrifice and attention. She may then ask herself whether that kind of focus would be worth the effort when much is so available for nothing to so many.

The second legacy is the understanding of the age, the considerations of the larger professional community that made centers necessary if certain alternative styles to educational inquiry were to be made realistic. She may go to her own considerations of the larger post-modern professional scene to find possibilities to carry alternative approaches and construct alternative principles of operation for inquiry and construction of knowledge in education. This may require, of course, different conceptions of the ways in which practice and its judgment can be addressed by those inside and outside of classrooms, or indeed what classrooms are in the post-modern world.

The third legacy is the robust success of centers when compared with individual work. This includes the impact of centers on professional thought, the range of styles of action within the limits of the principles followed, the variety of individuals connected to the centers over time, the robust identity with which centers were able to convey in their professional work, and in the distances between "followers" and "followed" when compared with individual stars of the profession at the same time. This might be the main reason why a post-modern maverick may consider the legacy of centers at all. Compared to what an individual can do to retain a robust "idea" that needs attention and work, she knows from the experience of centers that a small group of people will outperform her.

The fourth legacy is the financial performance of centers, the continuance of an institutional existence through thin and thinner financial times. How centers have survived the exigencies of funding patterns over the past 20 years can provide lessons for those interested in seeing their own visions of common effort enabled. Just exactly what does "healthy" mean for an organization formed at the fringes of a professional boundary, aimed at bringing the frontiers of analysis and reconsideration to untrained and uninitiated professionals (and public)? To what extent have centers surpassed what had previously been assigned to be "prerequisites" to the knowledge necessary to perform these analyses? I think the answers could be some of the more important legacies of centers for the post-modern bent on reaching around all the comers set before her.

The fifth legacy is the pattern of professional force, the significance that can be attributed to specific centers. Not just the total of the force would be the legacy, but when the force was felt, near the beginnings of the center, for example, or near the end. It is the patterns of professional influence over time that could be a legacy of centers, the ways in which such influence has waxed and waned, and perhaps evened out over the months and years. My guess is that most patterns of influence from centers will be similar, and thus could lead post-modern ambitions to new considerations of when to expect an idea to take form, and to end. Considering such legacy may provide some perspective on when to snuff the life of non-centers, and move to something entirely new.

The sixth legacy is the impact of the centers on the individuals who worked there. Certainly work in these centers was experienced differently than work in the more normal terrain of academic scholarship. What did this mean in professional and personal terms? What can we learn about the significance of non-independent, collective work on professionals trained primarily for autonomous and independent inquiry? Might there be similar challenges in our post-modern futures?

Although more can be made of the legacy of centers than my imagination has allowed, there is one caution to be made to anyone who wants to apply the legacy of centers to their own post-modern work. The reality is that the modern phenomenon of successful centers occurred at the twilight of the modern era. At about the same time that contemporary, post-modern art was being constructed, so were centers in educational inquiry. When post-modern critical philosophers were forming their thoughts, centers were hitting their stride. In other words, it not only took a long time for centers to be developed to respond to the educational challenges of the modern era, but they were successful near the cusp at the end of modernism and the beginning of post-modernism. Perhaps creative educational responses to this post-modern era will have to wait until it is nearly over, bordering on something else. But I hope not.

Case Study: The Importance of Multiple Takes

Jacquetta Hill
University of Illinois

It took me a good many years to think of an ethnographic study of culture as "case study" research. Perhaps it was because in that by-gone era "case study" was a pejorative term. I remember a renowned Comparative Educator of the University of Chicago judging "Case Studies in Educational Anthropology," a series edited by George and Louis Spindler, to be a set of samples of one so limited as an empirical data source as to be useless for analysis of education . . . quantitative analysis of course. To him, they amounted to little more than a very small sample. Bob Stake's work on case study and my own off and on participation in his course on "Case Study in Educational Evaluation" persuaded me that case study was the best data frame to put on the ethnographic/anthropological study of a school, a class room, a village, an organization, a household, a life history, a ceremony, etc., the common units of study for ethnographers of anthropological bent.

I

Always of course there is to be specified "*a case of what?. . .*" John Van Maanen, for example, holds that deciding what is to be counted as a unit of analysis is an interpretive issue of judgment and choice, where meanings rather than frequencies assume paramount significance. Indeed he advises ". . . think of qualitative method as procedures for counting to one. . ." (Van Maanen, 1988). That is *one* perspective that Bob also adopts in *The Art of Case Study Research,* along with a *second* and equally important discussion of the "collective case study," the sort of case study research that is of special interests herein. Bob gives us a thoughtful, useful typology of case studies: the intrinsic case, built on experiential research and aimed at understanding that can be conveyed in naturalistic ways to the interested reader of the account of the study. There is also, the instrumental case study that aims primarily to illuminate an issue and/or other cases. Instrumental cases at times come in sets, in both evaluations and in research to form collective case study. Bob and his students have worked through the debates of evaluation and research and the single case with their own work in mind, and offer us a variegated, but more orderly way to understand and explain what we are doing in case study work, ethnographic or otherwise.

Yet a case is actually never singular, as one must recognize as one acknowledges what Cuff (1993) dubbed "the problem of versions," a problem that Bob too handles with cogency and finesse. In the context of discussing generalization, naturalistic compared to "explicated (propositional) generalizations," one sees Bob's deeper, more variegated view of the case of one versus several cases in relation to one another (1995, p. 85).

In brief, with the unit spelled out, "case" for ethnographers gives entity status to an otherwise amorphous something. Even framed as a case of one, researched with richness, in a search for its own saliences, the "intrinsic" case can lend a hand or an insight, to local understanding. Still a case is handsomely enriched by the possibility of being placed in some relationship to another case framed as a similar kind of case in hot pursuit of a solution to a problem, or an "issue" as Bob prefers to label it. This is, in his paradigm, the instrumental case in a "collective study." There too a story hangs.

I would like to relate here the peculiar story of the *growth* of a "collective case study" out of a single case study, the formulation of an issue and a theory, and a dedicated, persistent researcher, John Ogbu. The studies of sites of more than 25 case studies ranging all over the world came to be placed in relation to one another in pursuit of understanding a puzzle: why some minority populations in the US achieved so poorly or failed in the schools, while others, also poor, denigrated and discriminated against, nevertheless succeeded so well. Not only did cases regarding the several minority, ethnic, and racially defined populations of the US collect around this issue and John Ogbu's effort to explain it, but as well, cases from Canada, the British Isles, the Netherlands, Germany, France, Israel, South Africa, Japan, Korea, and the West Indies. The issue was minority status and education, but the subterranean issue of race, shot, and surfaced again and again in the debates around the cases. One might call this queuing up of cases a line of research, rather like one witnesses in the physical or biological field, but seldom in the social sciences of education. However, the heart of this collective research was not experiment nor survey but commentary and comparison of collections of case studies bearing on an issue. Since this kind of growth of a "collective case-study" study is not ordinarily included in our discussions of case studies, although I don't think it is unique, I have thought it worthwhile to bring to the attention of seasoned veterans of case study research.

II

The intriguing queue up of case studies of minority education was initiated with a single case study by John Ogbu based on fieldwork carried out between 1968 and 1970 in the schools in Stockton California, that I want to sketch out here. Ogbu, interested in schooling and status mobility, in his single case study of a California neighborhood and its schools, drew on an array of both qualitative and quantitative data resources with all its rich complexity, to form a theory. That study and his vigorous discussions of extensions and modifications of it, set in motion chains of other case studies in reaction. Ogbu based his explanation for poor school achievement by African-Americans (as well as Mexican origin students) in the Stockton community on the social structural conditions of caste/class: the relatively poor school performance of African-Americans in Stockton, in spite of their wish to succeed, is rooted, he said, in the history of their involuntary incorporation into American society and the subsequent discriminatory treatment of them in a system of racial castes. Perhaps it is best to summarize his point of view directly:

> I would suggest that because of the amount of distrust that blacks have for whites and the schools controlled by the latter, it is difficult for black parents to teach their children successfully to accept, internalize and follow school rules of behavior made by whites, and it is difficult for black children to accept, internalize and follow such rules of behavior for academic achievement.

Low school performance was thus seen by the analyst Ogbu as an adaptation by Stockton African-Americans, as well as the

Hispanic origin population. This suggested the counterintuitive position that poor performance was adaptive.[1]

Heated debates broke out over the theory for which the case study was the database. It was perhaps both a source of data that refined a conceptual approach and an example of it. Some of that heat arose among the community of educational anthropologists, if I can call them a community, because his formulation challenged the idea that cultural difference was a primary factor in explaining success in school and occupational success thereafter. He also contested two earlier theories, two versions of difference as deficit: the biological deficit theory, specifically Arthur Jensen's version, and the cultural deficit theory, that a culture is deficient with respect to mainstream in its intellectual resources. The cultural difference concept was more sophisticated than the culture deficit theory, but nevertheless turned on difference. That difference Bourdiue had labeled "cultural capital" (1977) thereby winning new converts to the cultural difference position, understood in terms of social class. With that label cultural difference took on Marxian materialistic overtones, especially when cultural capital in schooling was linked to reproduction of the social order.

Ogbu's case study did not stand alone, but was surrounded by a complex schema of ideas about its relation to the effort to explain school achievement and why school achievement was or was not linked to occupational success and higher socioeconomic status. In its first stage Ogbu intended it as an intrinsic case study, but through the academic labyrinth of doctoral work it became an instrumental case study to illuminate an issue and problem, a puzzlement for which he offered a theoretical explanation. Subsequently a good many of those ideas were challenged and his explanations contested as well as supported in a growing array of case studies. Bob might characterize this case study research as instrumental case study: Each case study is instrumental to learning about issues and illuminating problems. But Bob might be unwilling to call the queue of case studies a "collective case study," probably because that kind must have ". . . important coordination between individual studies." The coordination among these cases is a looser sort, driven by a generalized explanation of an issue: Why do some minority populations do so well in US schools, while others do so poorly? Bob's view is that,"Case study seems a poor basis for generalization." Yet, generalization is part of case study in his view: "Generalizations about a case or a few cases in a particular situation might not be thought of as generalizations and may need some label such as *petite generalizations*, but they are generalizations that regularly occur all along the way in case study. . . . [Even] *grand generalizations* . . . can be modified by case study" (1995, p. 7).

---

[1] Ray McDermott has taken up the view of failure as adaptive, some years later phrased it as "achieving school failure." I found several of my African American students hotly contested his was an analytic version however.

At the same time many of the cases are clearly for situationally limited, intrinsic in purpose, fitting in many respects Stake and Trumbull's notion of intrinsic case study (1982). They are full of the intention to provide experiential understanding to sets of readers, including researchers, educators and academic professionals, so that their limited intuitive comprehension of "how it is" with this or that population, whose shoes they never have occupied, is experientially increased. And the best devices of qualitative research ". . . narratives to optimize the opportunity of the reader to gain an experiential understanding of the case"(Stake, 1995, p. 70) is expertly employed. It is an important variety of cases in the variegated landscape that these collections crisscross.

III

Ogbu began to modify his stance, not initially because of direct criticism and attack, as I see it; but because of the dramatic results of another case study--Greta Gibson's study of immigrant Punjabi South Asians in California who managed what she phrased as "accommodation without assimilation," and "additive acculturation." They were treated often also as racially different; but despite the denigration and derogatory experience, and the Punjabi youths' much resented resistance to American youth culture in the high school, they succeed academically and occupationally. This case, along with a growing body of challenging and complicating findings on Asian Immigrant populations, and then on Central American

Immigrant populations, led Ogbu to reformulate his theory around a second minority experience, the immigrant minority.

Following Ogbu's first study, the *Anthropology & Education Quarterly* published a set of four case studies under the title "Explaining the School Performance of Minority Students," along with five "framework" articles that drew on the four cases, arguing analytic points of view or versions. In response to the cases D'Mato, for example, after pointing out two different *versions* of the analysis--one by Ogbu and one Fred Erickson, remarks that their versions overlap (agree) in "looking at the school from the point of view of the society, rather than at the society from the point of view of the school." He goes on to point out that they miss entirely the **situated nature** of the point of view of the students. He illustrates the point with a case of his own, on teaching reading by two different approaches to the same students by two different teachers, in Kameamea School in Hawaii. In one class the children were chaotic, in the other, intently involved. For him, situated specificity and *attention to the problem of versions* was missing among the case studies assembled around the issue at that point.

Thirteen years after the publication of Ogbu's case study of Stockton, another collection of case studies on *Minority Status and Schooling: A Comparative Study of Immigrant and Involuntary Minorities* was published edited by Margaret Gibson and John Ogbu (1991). The "*involuntary*" or "*nonimmigrant minority*" refers not only to those who are native-born and minority, but to groups incorporated into the host society *involuntarily*, by means of colonization, conquest, or slavery, also assigned a subordinate status within the society. "*Immigrant minority*" refers to those populations which are actually immigrants and also to those whose ancestors were immigrants and who continue to maintain a separate minority-group identity. Immigrant minorities, like involuntary minorities, may also be denigrated and assigned subordinate position by the dominant group and suffer the consequences of prejudice and discrimination, but they consistently succeed educationally. It was a comparative study of cases that caught up with the consistent school success of voluntary immigrant minorities, even those that might be subjected to the denigrations of color bar, like the South Asian Punjabi's of agricultural areas in southern California. But the cases of Native American education looked very like the cases of many urban African-American education.

IV

In this growing collection of case studies, growing variety of "takes" in case studies on the questions of minority status, came to be highly reminiscent of the kind of case experience that Rand Spiro and his colleagues advocate for forming adequate knowledge structures for learning and comprehending very complex and hard to understand phenomena, (in its most extreme--complex phenomena in ill-structured domains). Here, of course, we are talking not of individual understanding alone, but rather a community of debating analysts, experts, developing **advanced knowledge**. Spiro and his colleagues work with knowledge of the body processes, such as that extraordinarily complex of knowledge of what is involved in heart failures (1988). Forming theories or models or schemas, for advanced knowledge of complex situations (as contrasted with the schemas formed of everyday routine situations [Gagne et al 1993, pp. 151-175]) it takes crisscrossing a variety of cases to build and develop advanced understand of very complex phenomena. It is the job and goal of the analyst to bring to the attention of colleagues the shortcomings of the complex knowledge paradigms they offer for presenting the case and its use for explaining a problem.

And in some ways cases and case studies can follow a steeper course of evolution for insight and understanding reminiscent of cognitive change like that found for forming complex knowledge structure in individual thinkers. But individual thinkers are operating as always, in a social context of discourse and use of their knowledge.[2]

Turning back to our story of cases, the collection of several case studies and "takes" criss-crossed a much wider world terrain of circumstances, turning up a more complicated pattern: Cultural difference counted differently in different societal political structural circumstances. (See Spiro, 1988, for a discussion of the cognitive "learning" significance of Wittgenstein's metaphor of criss-crossing cases.) For example, Koreans in Japan were treated as a racially different minority. Thus, an "involuntary minority," they perform poorly in school. However, the same population is an "immigrant minority" in the U. S. Here they do well in school and go on to higher education in significantly higher numbers than average. The case study offers what anthropologists have called a "naturally" controlled comparison. The success of first generation immigrants was a recurrent theme, as was the other theme of school failure of involuntary immigrant and indigenous minorities, in cases from New Zealand, Australia, the West Indies, and the British Isles.

---

[2] And according to Cuff (1993), individual thinkers in social contexts of everyday life address the problem of versions differently than scholars and researchers, as we said above, but unfortunately we can't here turn aside to examine the bearing of his ideas on this story of case study research.

Returning now to the historical line up of cases, it was only a year after the Gibson and Ogbu collection of ten case studies was published, that Suarez-Orozco edited five more case studies, for an AEQ special issue title, *Migration, Minority Status, and Education: European Dilemmas and Responses in the 1990s.* This time the cases were mainly from Europe where migration, minority status, and education pose dilemmas for the Europeans. Here the immigrations were recent, but the status of the immigrants on entry was quite different than the US. cases: Citizenship was not common and their political status was quite different . . . even for their European born children. The conditions of migration are different, and the structural factor that correlates with the differences in cases is being "non-European Economic Community" (Moroccans, Tunisians, and Turks) in origin or European Economic Community (Spaniards and Southern Italians) in origin. But with more passes through the cases in the discussions, a submerged factor with a very, very old history in European regions surfaced: Islamic and non-Islamic religious alliances of the immigrant populations.

The several case-study "takes" on the European settings turned up a "problem of versions." Much of the research takes the view of the hosts regarding immigrant success to be social, cultural, linguistic as well as economic assimilation into the host country as yardsticks by which to measure successful "adaptation." But the version of success that motivates many of the migrants in Europe is the ability to

RETURN to the home country, re-establish themselves and raise their social status THERE with the help of the money earned in foreign lands under difficulty circumstances. In the first generation the possibility of return *MAY* immunize these first generation "guest workers" and their children to the harsh denigration and treatment by host country schools of Belgium, Germany, the Netherlands, France, and British Isles. So the general immigrant willingness to play by the rules, that leads to success in schools of the European Union, is found in case after case in Western Europe.

V

So in the case study "*takes*," the matter of versions is ever present, and significantly multiplies the complexity of the enterprise.

In 1997 yet another set of case studies from Europe, and the Middle East on minorities, clearly complicates Ogbu's version of the theory of minority status and schooling, adding to the compelling necessity to mark the model by versions. Published as *Ethnicity and School Performance: Complicating the Immigrant/Involuntary Minority Typology*, and guest edited by Margaret Gibson, this collection shows once again the key significance of the students' version of conditions and purpose, and shows as well, their ability to form strategies of resistance within accommodations (a refined version of Gibson's "accommodation and 'additive' acculturation without assimilation"). Critical commentary took special note of student versions in overcoming the simplistic dichotomous reading of resistance and conformity. So it was observed in several cases that students may resist those acts seen as oppressive within their schools, but at the same time adopt strategies that lead to academic success, and subsequently to higher education and occupational and life style mobility.

So as case after case brought variety to the terrain of factors, conditions, people, and schools, as well as the issues, the crisscrossing in multiple takes was an essential contribution to the reformulations of older analytic versions and the theories based on them. But at the same time there never was a case that presented teachers' versions in a thorough going way. And its absence raises the question Why? This absence in all these many cases brings to mind the desirability of some teacher-as-researchers contributing their case studies to line of growth of case studies already here.

Another remarkable fact about this collection of case studies is that all the commentary and analysis has been discursive and linguistic in nature. Not once has any research subjected the cases as unit frequencies in any kind of quantitative summary or analysis. This whole enterprise has been qualitative and discursive. Is it then anti case study research to undertake an analysis of that sort? I think Bob Stake would argue, sensibly in my view, that it is not some violation of an implicit rule of case study research, nor even of qualitative case study to do a quantitative analysis in complement to the qualitative analysis. In fact I believe there is insight, understanding, and good information lost from view because no one has undertaken this complementary, alternative research procedure. In a different note, to my mind there is more. The cases, available in their qualitative richness, instead of only in the metonymic representation of sorted kinds of frequencies, are there for other analytic versions to be formulated across the whole set. So at the same time I would propose that, the more takes, the more fullsome the intellectual result, and the more firmly grounded the knowledge for use.

References

Bourdieu, P. (1977). *Outline of a theory of practice.* Cambridge, England: Cambrige University Press.

Cuff, E. C. (1993). *The problem of versions in everyday situations.* U. Press of America.

Spiro, R. J., et al. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structure domains. In *Tenth Annual Conference of the Cognitive Science Society* (pp. 375-383). Hillsdate, NJ: Erlbaum.

Stake, Robert. (1995). *The art of case study research.* Thousand Oaks, CA: Sage Publications.

Stake, Robert. (1994). Case study. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). Thousand Oaks, CA: Sage.

Stake, Robert & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science,* 7(1), 1-12.

Van Maanen, John. (1988). *Tales of the field: On writing ethnography.* Chicago: U. of Chicago Press.

Robert Stake and Our Business of Evaluation

Katherine Ryan
and John C. Ory
Office of Instructional Resources
University of Illinois at Urbana-Champaign

Introduction

Without question Robert Stake's contributions to program evaluation are well-known and highly-acclaimed. Work such as "The Countenance of Educational Evaluation" (1967), "Case Studies in Science Education" (Stake & Easley, 1979), and *The Art of Case Study Research* (1995) are representative of the depth and breadth of his work in this area. How his work has informed the theory and practice of evaluation is acknowledged in multiple ways including his recognition as one of the foundational theorists in the major text in evaluation theory (Shadish, Cook, & Leviton, 1991).

But as we gather this weekend and reflect on Stake's storied career we want to highlight his (possibly lesser known) contributions to the area of faculty evaluation in higher education. More specifically, we want to share the findings of an intrinsic case study to illustrate how Stake's evolving theory of teaching evaluation (Stake, 1971; Stake, 1987; Stake and Cisneros-Cohernour, 1998) informed a set of teaching evaluation practices in one particular institutional setting. Why did we choose an intrinsic case study? Using Stake's own words (Stake, 1995, p, 3),

> We are interested in it [the case], not because by studying it we learn about other cases or about some general problem, but because we need to learn about that particular case. We have an intrinsic interest in the case, and we may call our work "intrinsic case study."

The paper is organized in three sections. First, a brief overview of Stake's theory of evaluating teaching in higher education is presented as a background for the study. The case method and sources used in this study are then described. Finally, a summary of the findings is presented with a brief discussion of how Stake's work has and will impact the evaluation of teaching.

Overview of Stake's Theory of Teaching Evaluation in Higher Education

Stake has taken up the topic of the evaluation of teaching in higher education on several occasions. Unlike some other work in the area (Scriven, 1995), Stake's theory portrays the evaluation of teaching as a multi-level enterprise beyond the self-contained classroom and the lone individual instructor's responsibilities. For example, among other features, Stake's theory emphasizes the following:

- the evaluation of teaching should portray the complexity of teaching;

- the evaluation of teaching is inseparable from the evaluation of the institution;

- an instructor's contribution to instruction at the department level is an integral part of the evaluation of teaching;

- the evaluation of teaching includes the assessment of student learning; and

- recommends multiple sources, and naturalistic and quantitative methods for evaluating teaching.

To a certain extent, Stake's papers on the evaluation of teaching in higher education correspond to the evolution of his thinking about evaluation. Stake has three papers devoted to the evaluation of teaching in higher education. He presented his fundamental theory on the evaluation of teaching in higher education in "The Evaluation of Teaching: A Position Paper" (Stake, 1971). The theory was elaborated in 1987, "The Evaluation of Teaching on Campus." Most recently in 1998, using case material with Cisneros-Cohernour, Stake expands a critical feature of his theory, namely, that the evaluation of teaching should reach beyond the notion of the lone instructor in a single classroom. The extension provides an illustration of how to evaluate teaching using a community of practice approach that provides feedback via collective peer evaluations. (The peer evaluations are conducted by several faculty members.)

Below Stake's theory[1] is summarized within the following evaluation dimensions: purpose, context, evaluator's role, scope of the evaluation, methods or approaches, and use.

Purpose of teaching evaluation in higher education: In his 1987 paper Stake suggested there are at least four purposes for teacher evaluation: (1) Provide information for rewarding excellence and to improve areas of concern; (2) assist in the selection of best qualified candidates and the retention of currently qualified faculty; (3) assist in professional development for new and continuing faculty members; and (4) aid in understanding the institution at department or campus levels.

Context of teaching evaluation: Stake (1971) cited several factors to be studied co-terminously with teaching evaluation. To judge teaching appropriately, he suggests the values of factors such as institutional goals, school environments, administrative operations, curricular content, and student achievement should be considered as part of the context of the evaluation of teaching.

Evaluator roles in teaching evaluation: While Stake suggests "leaving instructors in charge" (p. 4, 1987), he sees both faculty members and administrators as the evaluators of teaching. Administrators are responsible for the encouragement and facilitation of teaching improvement. Faculty members are responsible for the improvement of their own instruction. [Stake is, to some extent, ambiguous on the relationship between the evaluation of teaching and rank and pay.]

Scope of teaching evaluation: Proposing that the landscape of the evaluation of teaching in higher education is beyond the single instructor in the classroom, Stake argues that the team contributions of faculty members should also be the focus of evaluation (Stake and Cisneros-Cohernour, 1998).

---

[1] This is, by no means, a complete discussion of Stake's theory. For example, this summary does not include his analysis of appropriate faculty comparison groups or his discussion of evaluation and the selection and placement of new and returning faculty.

Methods and Sources: Stake (1987) suggests naturalistic and quantitative methods for evaluating teaching. These include, for example, checklists of classroom conditions that promote learning, course content reviews by trained and experienced colleagues, student opinion polling services, classroom observations, and current and follow-up student achievement data. Evaluative observations are collected from multiple sources of data including students, peer faculty, and administrators. He emphasizes that findings from these sources represent different emphases and, as a consequence, are not likely to converge. For example, peers emphasize intellectual accomplishments and knowledge, which may not be the primary concern of students.

Use: Stake ties use to the purposes of the evaluation of teaching. However, when there are concerns to be addressed in teaching, he suggests using creative approaches. For example, he recommends that administrators should influence teaching through persuasion and/or providing additional resources (e.g., teaching assistants).

Method

The case study method was used for this study (Stake, 1995). Data were collected from unstructured interviews with informants that included former division heads and measurement and evaluation specialists who had worked in the institutional setting over the past thirty years (1967-1998). Interviews were conducted at the institutional setting with current and recent division heads. Heads and measurement and evaluation specialists from the distant past were interviewed by phone and electronic mail. A document analysis was performed on Stake's papers (1971, 1987, 1998) and institutional archival materials including inter-office and inter-institutional memos, internal research reports, and unpublished position papers. Findings from the interviews and document analysis were synthesized.

Findings

Mentorship

> "Bob Stake was my Mentor from afar."
> (Former Head of the Division of Measurement & Evaluation (M&E)[2] in the Office of Instructional Resources (OIR))

Robert Stake served as a mentor from afar for many former heads of Measurement and Evaluation. One head remembers hearing Stake speak on the evaluation of teaching at the University of Nebraska in the late 60s or early 70s. Another head actually took Stake's Case Study course as a graduate student at the University of Illinois, although her graduate training was primarily in quantitative methods. As an instructor in an introduction to evaluation theory course, she has students read his work as one of the foundational theorists in evaluation. A former M&E head considers his attendance at the CIRCE Brown Bag Seminars as a young research associate as a defining experience in his development as a practicing evaluator. Coming from a quantitative background when hired at OIR he reports "how much" those seminars broadened his perspective on evaluation.

How did this mentoring from afar inform the evaluation policy and practices of this campus evaluation unit? Following are some of Stake's "influences." A past head says that, historically, information from the evaluation of courses was considered "sacred." Student ratings were considered personnel evaluation information so they were never used for the purposes of program evaluations.[3] In addition, so students would not confuse the focus of the evaluation, program evaluation-type items were not administered as part of the ICES questionnaires. He cites Bob Stake's thinking on the uses of evaluation information as one source for OIR's long-term commitment to this policy.

---

[2] The name of the division was changed at least twice over the past 35 years. For the sake of clarity, the division will be referred to as M&E throughout the text. However, the authors acknowledge that previous division names include, for example, Measurement and Research Division (MARD) and Office of Instructional Research.

[3] In 1996, ICES data were aggregated across departments. Today they are one of the indicators in the Campus Profile. The Campus Profile is a system of centrally-provided indicators and unit-supplied information to be collected and reported annually for each department and aggregated at the college and campus levels.

The current head of M&E says that Stake's recommendation that both quantitative and naturalistic methods should be used for evaluating teaching is one of the primary reasons behind a new initiative for evaluating teaching. The division is experimenting with conducting focus groups with students enrolled in a course. "The focus groups can be highly contextualized and provide lots of information." The division is trying to formalize the approach and the results so it can be more easily used by the faculty in promotion or salary papers. The present head also described how they are planning to try-out the use of narrative reports and peer checklists in their trial evaluations of On-Line courses.

A past M&E division head suggests one benchmark he used to think about policy and practice was "How would RS (Robert Stake) come down on this issue?" In making decisions, this past head reports that he really tried to address what he thought would be Stake's concerns. These concerns always gave him pause. Nevertheless, these concerns were not always translated into practice. As this former head stated, "He (Stake) did not want student rating comparisons and was particularly concerned about routine student rating comparisons. Stake wanted no easy answers for determining excellence."

Student Ratings of Instruction

> "The old CEQ was spoiling a lot of good things that were happening for OIR on campus."
> (Former head of the Division of Instructional Development (DID))
>
> "The ICES cafeteria model was responsive."
> (Former head of M&E)
>
> "ICES was a more responsive student rating system."
> (Former head of DID)

> "The cafeteria approach of ICES tried to accommodate and to tailor evaluation to a specific context."
> (Former head of M&E)

Robert Stake's evaluation theory is known far and wide as "responsive evaluation." This evaluation approach focuses on program activities instead of goals, responds to stakeholders' concerns, and considers the different perspectives of stakeholders primary when making judgments about the programs (Stake, 1975). Two former heads of M&E suggest that the current Instructor and Course Evaluation System (ICES) was an attempt to be "responsive" to the evaluation concerns of the campus. What was OIR's take on the meaning of "responsive"? A past head of DID put it like this:

> I wanted to tailor my developmental work to the context and needs of each faculty member I worked with. I saw little value in asking the same questions or using the same methods each time I tried to help someone. To help me implement his form of responsive faculty evaluation I always looked to hire Bob Stake's graduate students; some became major influences in their own right.
> (Former head of DID)

How was this notion of "responsiveness" translated into teaching evaluation practice? The first student rating system used at UIUC was called the Course Evaluation Questionnaire or CEQ. The form consisted of a fixed set of items that appeared on each rating form. A former head of DID explained how difficult it was to use the same evaluation form with faculty in different departments, using different methods, in different settings (e.g., labs, studios). "I needed to use different items for lab and studio courses if I wanted to be responsive to each faculty member's needs," he said. The former head of DID and a graduate student of Stake allied with a new hire by M&E to begin planning a rating system that allowed faculty to select evaluation items from a catalog of items. Their early "secret" meetings led to formal committee work under the support and direction of the new head of M&E (brought to campus on the recommendation of none other than Bob Stake). Thus began the development of today's Instructor and Course Evaluation System, ICES. Or, what the former DID head calls, "a more responsive student rating system."

However, the notion of ICES as responsive teaching evaluation was not without complexities. All heads of M&E voiced uneasiness about students' ratings in general as a method for evaluating teaching and especially as a lone method. Furthermore, they were concerned about Stake's perspective on student ratings, in general and what Stake thought about ICES in particular. As one former head reported,

> I was very concerned about what Stake thought we were doing with ICES . . . that he think we were doing the very best we could with student ratings. I was so concerned about this I could not talk about student ratings in front of him. (Former head of M&E)

What was the impact of these concerns on evaluation policy and practice? As one former head said, "Bob emphasized the importance of portraying the complexity of teaching in the evaluation of teaching." All heads agree this emphasis persuaded them to think beyond student ratings in the evaluation of teaching in higher education.

Evaluation of Teaching

> "My respect for multiple methods and multiple sources in the evaluation of teaching grew out of my reading of Robert Stake's work and the CIRCE Brown Bags . . . "
> (Former head of M&E)

> "I tried to emphasize multiple methods and sources to address what I thought were his concerns about the use of student ratings in the evaluation of teaching."
> (Former head of M&E)

At least two heads of M&E link their commitment to multiple sources and methods in the evaluation of teaching in higher education to Robert Stake's thinking on methods and his concerns about student ratings. The commitment of M&E to multiple methods and sources in the evaluation of teaching is reflected in the historic and current scholarly work out of M&E (Braskamp, Brandenburg, & Ory, 1984; Braskamp & Ory, 1994; Ryan, 1997). Focusing on practice, Braskamp et al., laid out a framework for evaluating teaching in higher education that included multiple methods and sources. In an investigation of faculty use of teaching evaluation information, Ryan, 1997, found faculty were more likely to use information from students' comments (qualitative data) to improve their teaching.

> "Evaluation is everybody's business, but not everyone else's business."
> (Attributed to Robert Stake, p. II 8; Braskamp & Ory, 1994)

Stake (1971, 1987) suggests that the institutional context must be considered in the evaluation of teaching. Braskamp and Ory (1994) have transformed that notion considerably. They cite Robert Stake's comment in their book on assessing faculty work. In this book they spoke of using Stake's advice as a reminder that a balance must be maintained between the individual and institution in educational evaluation. Braskamp and Ory further suggest that when the balance tips in favor of the institution, a "climate of control, not commitment may be created on campus" (1994, p. 118).

The Future of the Evaluation of Teaching

> "Stake's latest paper with Cisneros-Cohernour fits our current philosophy and provides direction for the future."
> (Former head of M&E)

> "We must also recognize the differences across our departments. What is highly valued in one department may not be so in another. In Stake's most recent paper (with Cisneros-Cohernour) written for AERA in San Diego he speaks of the need to evaluate a faculty member on their contributions to their departmental community with an eye on particular accomplishments valued by the department."
> (Current head of M&E)

Former and current heads find the "community of practice" notion particularly suited for the current campus climate for the evaluation of teaching. Today's evaluation questions could include, "How can a faculty member's individual and collective teaching contribution be improved?" and "What is the merit or worth of a faculty member's contribution to the community of teaching: in the classroom, department, and the campus?" We should look to using multiple methods, sources, contexts, and criteria to answer these questions and to help faculty make a case for the quality of their teaching.

Closing

The results of our case study might lead one to believe that Bob Stake worked in OIR, or at least, met regularly with the staff. The truth of the matter is that is not the case. But as John and I thought of this weekend and of Bob's impact on so many people we realized how much his influence extended to our own work and that of OIR. And, this belief was obviously confirmed by the many comments solicited in our interviews with former employees. His influence was, and is, one of a mentor, of a colleague, of a leader in the field, of a respected critic, of a good friend.

References

Braskamp, L., Brandenburg, D. C., & Ory, J. (1984). *Evaluating teaching effectiveness: A practical guide.* Newbury Park, CA: Sage.

Braskamp, L., & Ory, J.C. (1994). *Assessing faculty work: Enhancing individual and institutional performance.* San Francisco, CA: Jossey-Bass Publishers.

Ryan, K. (1997, April). *Making student ratings comprehensible to faculty: A review of alternative reporting approaches.* (Faculty Evaluation SIG). Paper presented at the American Educational Research Association, Chicago, IL.

Scriven, M. (1987). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation, 21,* 119-129.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage Publications.

Stake, R. (1967). *The evaluation of teaching: A position paper.* Unpublished manuscript, University of Illinois at Urbana.

Stake, R. (1975). An interview with Robert Stake on responsive evaluation. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 33-38). Columbus, OH: Merrill.

Stake, R. (1987*). The evaluation of teaching on campus.* Unpublished manuscript, University of Illinois at Urbana.

Stake, R. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R., & Cisneros-Cohernour, E. J. (1998, April). *Evaluation of college teaching in a community of practice.* Paper to be presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Haluk Soydan

Maria Saez & Antonio Carretero

Delwyn Harnish, Philip Zodhiates,
& Najmuddin Shaik

Social Work Evaluation in Sweden and Robert Stake

Haluk Soydan
Gothenburg University, Gothenburg, Sweden.

The Background

The aim of this paper is to describe the background and the mission of the Center for Evaluation of Social Services (CUS) at the National Board of Health and Welfare. Doing this is also to undertake a short journey into the state of affairs of the Swedish social work practice research in general, and evaluation of social work practice in particular, in Sweden.

The idea of using evaluation in public administration within the Swedish welfare state is not new one. As elsewhere, evaluation in Sweden has emerged as a part of welfare state activities. Its early emergence was to be seen in the sector of public school education. Massive school reforms that were initiated during the 1950s. In the beginning of 1960s the comprehensive school system was adopted and an administrative body with evaluative function was set up. Bob Stakes' early Swedish contacts were initiated, as I later learned, as a consequence of this development. Evaluation activities at the departments of education were developed as theory-oriented, not to be mixed with theory-driven, social research at the service of the national school administration system. Theory-oriented evaluation did not aim to test hypotheses as the theory-driven evaluation, but to measure program effects with theoretical sensitivity. This development resulted in rational, instrumental evaluation research activities related to government reform work, a genuine example of social engineering skills.

If evaluation is defined as "an ex post mechanism for the systematic mapping and assessment of public policies and programs, their implications, outputs and outcomes for the purpose of effecting future decisions" (Vedung 1995, 72) its emergence is very recent in the Swedish social work practice. It is not before 1990s that evaluation in its modern formation was introduced to the thinking of social work research and practice, although the idea of research-based social work practice is not new and was explicitly discussed by the State Commission on Social issues in 1974 which prepared the Social Services Act of 1982, and the reformation of social services.

The State Commission stressed the importance of social work research in development of social work practice. This concern was based on interest in the generation of new and fruitful knowledge for social work practice as well as on understanding of the importance of professional and client perspectives on the outcomes of interventions. The Commission pointed also out the importance of systematic assessment of local experiences, but did not so much stress the potential role of social work practice in development of sustainable knowledge for practice and for client outcomes. As a part of the same reform other measures were taken: most important was the establishment of social work as a research discipline with own research professorships, and as well as the establishment of a new research fund to support practice relevant social research. Typically, an important remit of the first chair in social work, as well as of those to come, was to produce research relevant to social work practice by studying social problems and formulating solutions. Social work research was also expected to take into consideration needs of social work education.

What happened, then, during the last two decades? Basically, there has been a lack of applied, practice relevant research which might promote more evidence-based social work practice. Instead, social work researchers gave more attention to social policy matters and macro-oriented research as well as general social criticism. A review of the research funded by the Swedish council for social research between 1991 and 1994 (SFR, 1994) shows that only about 15 percent of approximately 300 funded projects on social work include some sort of evaluation. Client-oriented projects in social work were not more than 10 percent of the total number of projects. Only in the area of alcohol and drug abuse more than 40 percent of the projects included research on effects of various types of treatment methods (Tengvald, 1995).

A review (Dellgran and Höjer, 1996) of doctoral dissertations prepared at the departments of social work showed that very little attention was given to evaluation of methods of social work. Furthermore, as social work research became integrated into the academic system questions about the role of the discipline were raised (Bäck-Wicklund, 1993). Was social work research to play an autonomous role in order to raise and investigate critical issues of social work practice, or was it to serve social work practice by being instrumental to the issues formulated by social work practice? While this question is very much in the agenda and being debated by researchers and practitioners, systematic research on the value of social work for clients has not been given priority by social work academics. In this sense the disappointment of social work practitioners with reluctance of academics is well-known.

Evidence-based knowledge in the service of social work practice might also be produced within the social services by practitioners. The State Commission on Social Issues of 1974 (SOU, 1974: 39) had already for more than twenty years ago expressed the necessity of continuous follow-ups and evaluation within social work agencies which in Sweden are organized by the municipal authorities. Specially, it was stressed that experiences of social workers should be systematized and best practices identified and disseminated to other sites of social work. It seems that very little has been achieved in this respect. Systematic knowledge on locally based follow-up studies and evaluations is very limited. Furthermore, it has not been usual to study outcomes and effects of social work interventions in Sweden. There has so far been carried out only one major study of evaluation research utilization within social services in Sweden (Nilsson & Sunesson, 1988, 1993a, 1993b). Also the recent State commission (SOU, 1994) on social services stresses the crucial role of evaluation in the local social services. Although there is no systematic picture of evaluation activities of local agencies, although a base-line study of their character and frequency is now carried out by CUS.

The Center for Evaluation of Social Services

The Center is an institute for evaluation research on personal social services at the National Board of Health and Welfare in Stockholm and was established in 1993. The raison d'être of the Center was based on the following proposition arguing that the social work profession "lacks systematic empirical validation of its practice strategies. Ongoing evaluation of social work interventions seems to be a desperate need all over the world" (Hokenstad et al., 1993: 187). The establishment of the Center was a compensatory measure where university departments of social work, other research centers as well as social work agencies have been reluctant to study outcomes of social work interventions.

The center operates five major evaluation research programs and a best practices program. Three programs are set up to observe three traditional sectors of the Swedish social services, namely, child and adolescent care and protection, treatment of drug and alcohol abuse, and social welfare and economic assistance. The two other research programs are cross-sectorial. The program for theory and practice of evaluation research is the strategic program of the Center and aims to develop better conceptualization and implementation of evaluation research in the field of social work practice. The program for migration, ethnicity and social work is also cross-sectorial in its character and is motivated by the fact that Sweden's demography has become multi-ethnic during the recent decades and a growing number of clients within the social services have diverse ethnic backgrounds. The program for best practices aims to identify good examples of social work methods and interventions, and to disseminate those experiences. Evaluation research programs include not only projects of research reviews and systematization of the state of the art in respective fields of activity, but also empirical studies, often comparative, longitudinal and quasi-experimental. Furthermore, the Center is busy with conferences, workshops and lectures in order to better reach social work practitioners as well as the research community. The general and long-term aim of the Center is then to contribute to a well-founded professional discourse in social work, characterized by theoretically sustainable and empirically substantiated studies of outcomes of social work interventions.

Is it possible to characterize development of the Center in a global context?

I would say, yes it is. Given the fact that the Center is not a university research department, neither an institute based in local social services context, the question about its nature becomes crucial. Although it is too early to predict the long-term development of the Center some tendencies can already be explored. I believe such tendencies might be understood in terms of the concept of "new mode knowledge production."

Michael Gibbons and his colleagues (1994) draw our attention to the emergence and development of a new type of knowledge production, based on empirical observations of the researchers in a global context. This group of researchers argue that a new mode of knowledge production has emerged and is developing as opposed to the traditional disciplinary production of knowledge. The new mode of production grow up to avoid shortcomings of the traditional knowledge production, in particular in terms of imperfect relationship between knowledge production and knowledge utilization.

The new mode of knowledge production is recognized by problem orientation, transdisciplinarity, organizational diversity, social accountability, reflexivity and quality control. Problem orientation has to do with the primary and central interest in problem solving and in organizing activities around given applications rather than necessarily following paradigmatic rules of a given discipline. The purpose of the research is then to solve given problems and not necessarily to satisfy disciplinary methodologies. By being strongly problem driven the new of mode of knowledge production transcends disciplinary borders and creates conditions for action and application.

Transdisciplinarity involves four separate but interlocked aspects. First, it strives to develop framework for problem solving in the context of application. As known, in the traditional mode, the knowledge production and the knowledge application usually belong to different contexts. Second, transdisciplinarity does not necessarily aim to generate disciplinary knowledge, even if solution to problem solving involves both empirical and theoretical elements. Third, the new mode presupposes continuous communication between researchers and stakeholders in order to secure efficient transfer of results. Fourth, transdisciplinarity means contextuality, that is knowledge production and application is a single context. The knowledge then might be the basis of problem solving in other contexts even if there is no simple guarantee for this type of generalization.

The new mode of knowledge production is characterized by organizational diversity. It might take place in independent research centers, government agencies, industrial laboratories, think-thanks, consultancies as well as in university settings. The traditional knowledge production is almost exclusively university based. Modern communications means and globalization of interaction arenas for scientists have been a necessary infrastructure for linkages and interaction between various types of research sites.

There is growing public concern and civic activities about the advances of science and technology because of the awareness among people in general of how research results may affect public interests. Consequently, public in general and organized interest groups in particular demand accountability and thus affect knowledge production. The new mode of production is sensitive to social accountability in terms of the definition of the problems, the setting of research priorities, and the interpretation and dissemination of research results. The quality control in the new mode of knowledge production is composite involving not only peer review judgments but also by taking into consideration criteria such as the market competitiveness of the solution, social acceptability and legitimacy, and cost efficiency.

Although there is not a official declaration in which the Center is characterized by traits of what has come be called the new mode of knowledge production, internal policy discussions and the way of setting up and running research projects at the Center resemble more and more the model of the new mode of knowledge production.

Stake and the Center

Professor Robert Stake's Swedish connections are rich. He has been involved in discussions of evaluation research in circles of Sweden's pioneering evaluation researchers at education departments. His keynote presentation on new trends in evaluation in 1973 at the school of education in

Gothenburg is a good example of his early contributions to the Swedish discourse (Stake 1973).

Furthermore, it is an honor for the writer of this contribution to the Stake Symposium to call attention to the fact that Robert Stake is a honorary doctor of the Faculty of Social Sciences at Uppsala University, where the present writer has accomplished his doctorate in sociology, and as pure coincidence, in close cooperation with a prominent University of Illinois psychologist, the late professor Charles E. Osgood.

Having Robert Stake's early Swedish engagement as well as his outstanding writing as a backdrop, it was natural for the Center to call for his participation in an international conference on evaluation as a tool in the development of social work discourse that took place near Stockholm in April 1997 (for the proceedings of the conference refer to the special issue of the Scandinavian Journal of Social Welfare 1998, nr 2). Robert Stake has made major contribution to the success of the conference not only with a paper prepared together with Linda Mabry but also by commenting on other contributions. His social charisma has attracted other prominent evaluation researchers to the conference as well as facilitated socializing process among conference participants. The paper, "Ethics in Program Evaluation" (Stake and Mabry 1998) discussed ethical issues in social work evaluation with great authority, based on many years' experience of practicing and conducting evaluation research. The understanding of ethical issues in social work proposed by the authors has been very illuminating for the continued activities of the Center. Robert Stake understands ethics as "the sum of human aspiration for: honor in personal endeavor, respect in dealings with one another, and fairness the collective treatment of others." Since social work will always contain dilemmas of difficult choices, the argument forwarded was that ethics will mean balancing competing principles, and not so much following ethical codes of pertinent institutions.

We hope to continue profiting from his knowledge and wisdom in our work to develop achievements of the Center.

References

Bäck-Wicklund, M. (1993). *Vetenskap och Politik i Socialt Arbete.* Göteborg: Institutionen för socialt arbete, Göteborgs univeersitet.

Dellgran, P. & Höjer, S. (1996). Hårda fakta om mjuk data. *Nordisk Sosialt Arbeid,* 3, 204-216.

Hokenstad, M., Khinduka, S. K. & Midgley, J. (1993). *Profiles in International Social Work.* Washington DC: National Association of Social Workers Press.

Gibbons, M. et al. (1994). *The New Production of Knowledge.* London: Sage

Nilsson, K. & Sunesson, S. (1988). *Konflikt, Kontroll, Expertis. Att använda Social Forskning.* Lund: Arkiv.

Nilsson, K. & Sunesson, S. (1993a). Conflict or control: Research utilization strategies as power techniques. *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization, 6*(2), 23 - 36.

Nilsson, K. & Sunesson, S. (1993b). Strategies and tactics: Utilization of research in three policy sector contexts. *The Journal of Applied Behavioral Science, 29*(3), 366 - 383.

SFR (1994). *Projektkataloger: Rörande Beviljade Forskningsanslag 1991 - 94; Rörande Barn och Familj, Hadikapp, Missbruk, Socialtjänst, Äldre.* Stockholm.

SOU (1974). *Socialvården - Mål och Medel.* Principbetänkande av Socialutredningen, 1974: 39. Stockholm.

SOU (1994). *Ny Socialtjänstlag.* Huvudbetänkande av Socialtjänstkommittén. 1994: 139. Stockholm.

Scandinavian Journal of Social Welfare. (1998). Special Issue: Evaluation Research and Social Work, 7(2).

Stake, R. (1973, October). *Program evaluation, particularly responsive evaluation.* Keynote presentation at the conference on "New Trends in Evaluation" at the Institute of Education at Göteborg University.

Stake, R. & Mabry, L. (1998). Ethics in program evaluation. *Scandinavian Journal of Social Welfare, 7*(2), 99 - 109.

Tengvald, K. (1995). *Kompetens och Kunskapsutveckling.* Bilaga 7 i SOU 1995:58.

Vedung, E. (1995). *Five observations on evaluation in Sweden.* Uppsala: Uppsala University, Department of Government.

Case Studies Approach in the Negotiating Evaluation Model

María J. Sáez Brezmes and Antonio J. Carretero
Valladolid University, Spain

When I presented this paper, I asked for feedback from colleagues attending the symposium. I got two questions. We are going to comment on those two questions in the appropriate section here.

Setting the scene

We would like to begin in an unorthodox way by telling how I (Sáez) became involved in evaluation and in the particular kind of evaluation in which I am interested. My background is biology, more precisely, biochemistry and cell biology, a field of research far distant from educational evaluation, but my interest in education brought these two disciplines together in my career. I later came to realize that quite a few influential people doing evaluation in Spain have backgrounds in science, mainly in physics or biology.

My concern in evaluation comes from an interest in gaining a deeper understanding of social change, particularly the impact of educational programs and policies, current problems of social acceptance and main issues in the development of evaluation in Western societies.

I started to study evaluation in East Anglia's Center for Applied Research in Education, with Barry MacDonald who was helping me give the first evaluation courses in Spain. Barry introduced me to how to set up and carry out evaluations and to their political nature. Ten years ago, Ernie House, who spent a few months at my university, helped me to understand the role of evaluation in the Spanish context. The definition of an evaluation model in that context led me to focus my study on the history and methodological problems of evaluation, as well as on credibility.

I was the first to invite Bob Stake to Spain because of the importance of case studies as a method and methodology for social sciences. Bob came to teach a seminar in case studies in the course I was giving–at one of the Madrid Universitie--

for policy-makers and other professionals who need to evaluate the implementation of social programs. I invited Bob again to participate in training policy-makers working for the then just created National Institute of Evaluation and Quality, designed to evaluate the educational system up to university level. He invited me to spend time in CIRCE with him, but the work at my university only allowed me to stay for a short period. Bob introduced me to the American Evaluation Association. Whenever I asked him for help in my professional development, the answer was positive. It was support and a sponsor that many would like to have had themselves.

I might say that Stake's view of case studies has deeply influenced me. He has been an important influence on the Spanish context. I translated some of his first papers into Spanish. Even though case studies are not yet a method widely used for evaluation in my context, academically speaking, interest is being developed, recently reinforced by the publication of Bob's book "The Art of Case Study Research" in Castellano.

For someone like me, with a strong science research background, trained in the experimental method of natural sciences, the research method for studying social facts was a relevant issue in my preparation as evaluator. That was, in fact, my main concern when I moved academically to a Faculty of Education. I was convinced that in order to understand educational issues, answer research questions posed in the field by practitioners and represent the complexity of social life in change, as was going on in my country, a methodology which can approach the phenomena in its complexity was needed.

The role of science in today's society is becoming more and more a relevant issue. As C. P. Snow (1969) remarked, there are two coexisting cultures in society today. Communication between these two is difficult, partly because knowledge is focused on what each of them does. For many years physics focused on simple phenomena and the social sciences focused on phenomena considered complex. Our perception today about these two types of phenomena is changing, the physical-natural phenomena even at macroscopic or microscopic level can no longer be understood as simple.

In the 19th century, time was introduced into the conceptual frame of the classic sciences, but past and future were basically understood as the one and the same. It was the Darwinism and the thermodynamic principles based on evolutionary paradigms which put time's arrow in the heart of scientific knowledge due to "irreversibility," fundamental for both of them. Within biology, the generation of new structures allows time's arrow to be understood from a constructivist perspective. The time paradox had deep consequences for modern science.

I. Prigogine, Nobel Chemistry Prize winner in 1977, in his book "Le leggi del caos" (1993), maintained that formulation of the traditional natural laws was contrary to the fundamental laws and the phenomenological descriptions of nature, basically because they did not include time. In the classic perspective, all of the laws of nature assumed a determinist and reversible description of time. The chaos theory inevitably introduced the concepts of probability and irreversibility, i.e. a new fundamental description of nature should be assumed along the lines of chaos theory.

Margulis and Sagan (1996) recently approached the same issue in a book with the same title as Schröedinger's, where she pointed out that:

> ". . . the maintenance of the body existence and the self reproduction are in the heart of the nature of what the life is."

Knowledge of ourselves as organisms consists of establishing a few basic principles applying to all living organisms. Darwin showed us that living organisms have a unique common ancestor. Margulis and Sagan not only tells us that this common ancestor is the bacteria but that the predecessor of the ordinary cell of our bodies is an amalgamate of bacteria strains. Margulis and Sagan emphasizes the importance of "symbiogenesis" in evolution, as a mechanism generating new living structures. This mechanism probably has been much more important than Darwin followers could ever have imagined, immersed in a tradition where competition was more relevant than cooperation in the evolutionary process.

V. Verdansky (1863-1945) described life as "living material." Fifteen years later, Lavelock described the surface of the earth, including the rocks and the air, as living organisms. Life is a process which can only be understood in the context of the earth planet. It is a process which rides matter like a wave, a chaos where a mix of chemical reactions produced a mammal brain 80 millions years ago, which in its current shape writes love letters and uses computers to calculate the temperature of matter in the origins of the universe.

Many biologists, experts in biochemistry, support the idea that we will understand life when we understand the different behaviors of living organisms in their own contexts for surviving. Uniqueness and diversity are two sides of the coin, perceived as a complex relationship between both.

I assume that the uniqueness of the case and the facts described in early fieldwork channels focuses the later collection of data. These should illustrate the relationship between the agents and the facts, and serve the purpose of thoroughly describing the everyday experiences that characterize the relationship. As Barry MacDonald (1987) outlined, from the data thus collected, it is realistic to expect particular facts proper for the case, but also data related to general and universal features. The categorization and processing of data would allow distinguishing the one from the other: It should be possible to discern those which are essentially idiosyncratic or specifically related to contexts to which they belong.

The first audience question was from David Jenness about the difficult connection between the complexity of life in biological terms and the negotiation model that I was proposing for evaluation.

Let me explain first which type of evaluation I am immersed in. We are evaluating the introduction of biotechnology in secondary schools for which EIBE has prepared activities and units for all of the European Union countries. The teachers (participating voluntarily) were to decide, at least in Spain, what and how they are going to implement it in their classrooms and schools. Issues about the impact of science in society and the societal use of biotechnology products is being discussed publicly among philosophers, biologists, historians, sociologists, etc. At the same time in the last decade, several studies have focused on the "public understanding of science" (Durant, Driver, etc.).

Reflections about biotechnology appeal to notions of "risk society" and "reflexive modernity," as defined by Beck

(1998), and suggests that "to advance along the pathways of reflexive modernization is to cause skepticism to spread and to reach the very foundations of scientific activity and the risks involved in the latter, as a result of which science itself becomes at once generalized and demystified." Whilst this generalization of science entails that society and its institutions inwardly regard scientific activity as being something unavoidable, at a time when the latter is undergoing privatization and is becoming more and more an economic undertaking, the demystification process implies that the role played by scientists and experts in general is constantly subjected to public scrutiny. To quote E. Muñoz (March's Public Seminar about "Modern and Postmodern Science"), the immediate consequence of this is that "experts have been placed under suspicion, a situation that is vividly illustrated by the field of biotechnology." From this point of view, technoscience has become a potential generator of social conflicts, owing largely to the lack of confidence placed in the experts by citizens. It is that there is "a need to create instruments that will enable us to negotiate and to reach consensus" on the values, priorities and risks involved in scientific research and its social consequences. In our opinion, it is precisely here that evaluation can and must fulfill a vital role, by bringing into action a model of negotiatory evaluation that extends evaluation as a political strategy to resolve conflicts and to further democratic dialogue (Saez & Carretero, 1995).

For Bob Stake (1995), researchers using case-studies as their *research strategy* adopt different roles. According to this perspective I was able to identify among others, three main aspects of Stake's work which were the most relevant to my own work in the Spanish context.

The singularity of the case and the types of phenomena that should be studied. The emphasis on uniqueness does not mean that only topics concerning small populations could be addressed. The 13 evaluation case studies about American science education in 1976 completed with Easley (1979) provided the best example of what can be done with case studies in this respect.

"Collective case study" is defined by Stake (1995) as the study conducted on individual cases whose common features are unknown before they are selected, even though they are chosen because, thanks to them, researchers expect to gain some insight into the phenomenon they are studying. Case-study reports based in multi-case studies, using several cases as empirical support for building the case and using cross-examination through comparison and contrast of facts and evidence in the cases under consideration provide the frame for using case studies for different types and sizes.

Data processing from different cases was made by contrast rather than by a comparative analysis. Contrast analysis provided the basis for understanding the process of innovation and change. Comparing cases among themselves and contrasting the relevant data collected at several of them provides a picture in different contexts. It is the contrast analysis that contributed most towards the holistic view of the field of study and figured out a pattern for the global perception of the object under scrutiny, allowing us to see the gradual development of change.

Which is the case? The case as a social construction. Writing case-studies with a naturalist methodology means that the educational facts are described in their context. Stake's idea of following certain "footprints" in the field of study so as to define the case as research goes along these lines. For him cases are theoretical constructs created by researchers in the course of their work, i.e. a "case" does not exist until the authors have created it.

I am building the case and developing the study's argument through comparison and iterative construct, what I call *progressive discernment* and because it is an exercise of cooperative construct shared by all collaborators in the evaluation (Well, 1995), it creates the "negotiating" commitment of those involved, whether directly or indirectly, in the making of decisions relative to the object of study.

Because I make a difference between the case and the study, the progressive discernment strategy really means producing feasible answers for the questions formulated in the course of intensive focalizing, answers that are explicative of the case's complexity. This is largely the result of opening up our scope to encompass the educational context of innovations analyzed in particular case-studies.

From a conceptual perspective, we use the term *progressive discernment* to refer to a research process whereby we are allowed to detach ourselves from an area of concern we have closely scrutinized to place it in its right context again. The exercise can be likened to the observation of a mitochondrial crest through an electronic microscope: we need powerful magnifying lenses. We would begin by identifying that crest as part of the mitochondrion it belongs to, and then as part of the cell in which it is. In order to affect that shift of perspective we need to quickly change to a lens of lesser magnifying power. Processes occurring on the mitochondrial crest can then attain significance within the particular cell-type under consideration and in terms of cellular respiration within euchariotic cells. The way of dealing with this type of cell and its assignment to the general class of euchomatin cells will be supported by as much data as is necessary, depending on what it is that we want to stress and which audiences we are addressing.

In fact, as Helen Simons (1995) pointed out, the last step of writing the case is tuning up with audiences, refining the argument, making sure that for each discussion of a key issue all positions are sufficiently represented and that, although defining the case necessarily entails a reduction of the data, such limitation will be compensated by references for further data consultation.

The Audiences and the Evaluators' Role

Stake's frame of the responsive model and the formative and summative concepts outline the relevance of audiences in evaluation. Cronbach (1982) and House (1993) agree that, to a certain extent, evaluation theory deals with political interactions and with the selection of facts summated under accountability, i.e. that evaluation validity cannot be separated from its political and social circumstances and considerations.

Are case-study evaluations useful in bringing about the involvement of audiences? If evaluation proves capable of providing such audiences with a deeper insight into the type of problems and developments which they are involved in, the answer will be affirmative. My aim has been to show how case studies are useful in both describing, in an accurate, empirical way, the phenomena under scrutiny and reasonably formulating the directions of change.

But, by analyzing this evaluation methodology, we shall be able to provide a heuristic answer to the question, "What constitutes the case in an evaluation?" This will in turn allow us to define the case as a social (and, therefore, collective) construction, proposing negotiation as a model for evaluative action in order to both identify main issues and disseminate the information collected.

When we reach a settlement with participants about their rights, their degree of participation in the process decisions, the publicity and dissemination given to the information supplied by them, we are actually proposing a series of ethical rules to be followed. These ethical principles become instrumental in trying to contrast our data with those of participants and other points of view.

Evaluation deals with information-related phenomena, acting as a means of mediation among different pressure groups. The value transfer process that is carried out during evaluation consists of dissemination among the various audiences involved, of distinct areas of information on the program as a whole (Saez & Carretero, 1998). Each and every one of the audiences has to feel that their own interests have not been neglected in the production of the information, while at the same time they are expected to adjust their interests, in light of the information received, from the other pressure groups. The need to be able to offer an alternative vision of the program, one which gives rise to a new understanding of the latter, is the factor that brings the different participants to adjust their value judgments to the findings afforded by evaluation. Negotiation strategies play a vital role in the search for agreement among agents who have come into conflict, the basis of said strategies being the information generated by evaluation. Indeed, the negotiation and the resulting recognition of discrepancies constitutes the postmodern ethos of evaluation. Although evaluation is not a guarantee for negotiation, negotiation forms an inherent part of the evaluation process, of its *modus operandi*, as an efficient means of reconsidering problems, thus ensuring the effectiveness of negotiation.

According to Stake (1995), validation of the data within a naturalist approach is called "triangulation". Even though I do not disagree with this idea, we prefer to use a wider concept like negotiation and raise this concept to the status of an evaluation model, in view of its fruitful methodological implications. In so far as we assume that negotiation is a process (within the larger process of intervention evolved by field evaluation), we must also take for granted that it becomes a means of validating whatever can "represent" the reality of the object of evaluation and the social relations and social actions that form it, especially if we consider the limited character of the data collected. The evaluator, then, plays the role of mediator--one who distributes information among the different groups involved. The latter entertain legitimate albeit different interests in the program, and the evaluator can attempt to build effective communication channels on the basis of a better understanding of the social situations of participants and to improve the efficiency of organizations (Carretero, 1995).

The second comment rather than a question came from Barry MacDonald. He expressed dismay that I was spending so much time talking about science rather than about the issues of negotiation in evaluation?

The final aim of evaluation is to express an opinion regarding the merit and the value of what is being evaluated. As a result of the increasingly important role played by audiences in evaluations, evaluative judgments have become more and more relativist and contextual. This means that the judgments are generally subjected to public scrutiny by the audiences taking part in the evaluation. On judging what he observes, the evaluator does not believe his judgment to be the only one possible, but rather considers it to probably be the most consistent, the best structured and the best founded. The consistency, structure and foundation that underlie evaluative judgment is a fruit of the evaluation process itself: the gathering and analysis of data places great emphasis on description of the situations encountered and encourages a diversity of opinions, the aim being to offer a comprehensive, plural interpretation of the complexity of what is being evaluated. The fact that evaluative judgments are deemed worthy of consideration is to some extent due precisely to the insight that audiences obtain from evaluation. Only on the basis of the alternative understanding that evaluation provides do the opinions expressed by the latter come to be admitted, considered, debated and carefully weighed up.

It is clear that the relativism of evaluative judgments stands in direct proportion to the demystification of science promoted by present-day post-modern movements. This in turn leads us to review the concept of objectivity put forward by Scriven (1997), which is to be understood as the ability of the evaluator to distance himself from what he is evaluating and therefore to afford an unbiased vision of the situation in hand. If we take evaluative judgment to be inherently relativist in nature and consider its acceptance by clients and audiences to depend on the insight it offers to the latter, the question arises as to whether or not evaluation should formulate recommendations and indeed whether evaluators should give their clients advice on which strategies to take and which paths to follow regarding the subject of the evaluation. If the answer were "yes," then evaluation would run the risk of being absorbed by companies offering consultancy and advisory services and would become just another instrument supporting the establishment of recommendations. Evaluation, however, responds to a different kind of demand, *viz.* that of those people who want to know in greater detail both the way a given program works and the current state of affairs surrounding it, and who wish to learn where the good points of a program lie and to calibrate its efficiency, effectiveness and suitability–in short, its value and merit--so as to be able to decide whether it should continue or be changed. In order for this to happen, it is essential that judgments be expressed. Inevitably complex, the latter provide possible action frameworks for those who have to take action and come to a decision regarding the program in question.

All recommendations presuppose a judgment, but very often this judgment is hidden in the recommendation itself. On the other hand, judgments do not necessarily entail unanimous recommendations, since their evaluative nature renders them transparent and open to public questioning, thus opening up the debate on the various alternative actions available. Political and social decision-making processes need such judgments–more than they need the technical recommendations which, limited in scope, are offered by advisory services. The provision of expert advice is a technical process well-suited to short-term decisions. Evaluation is a political process which facilitates strategic medium and long-term decision making. There is no doubt that evaluation can give rise to recommendations of a strategic nature, but such recommendations are always based on a plurality of interests and democratic dialogue, and never on the implicit or explicit presentation of the client's urgent needs.

Whether this negotiation-based model of evaluation is capable of giving a satisfactory response to institutional intervention and institutional learning will essentially depend on the explicative power of its political and organizational analysis. In other words, it is the theoretical and empirical support of evaluation that will ultimately generate a call for negotiation as the formula that regulates the assessment of the program's efficiency and usefulness as well as the decisions to be taken in this regard (Saez & Carretero, 1995).

References

Beck, U. (1998). *La sociedad del riesgo.* Paidós.

Carretero, A. J. (1995). La negociación en evaluación. En M. J. Sáez (ed.), *Conceptualizando la evaluación en España* (pp. 27-36). ICE Univ. Alcalá de Henares.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco, CA: Jossey-Bass.

Driver, R. (1996). *Young people images of science.* Open University Press.

Durant, J. (1993). *Science and culture in Europe.* London: Science Museum Publisher.

House, R. E. (1993). *Professional evaluation: Social impact and political consequences.* Beverly Hills, CA: Sage.

Margulis, L., & Sagan, D. (1995). *What is the life?* New York: Nevraumont Publishing Company.

Lavelock, J. E. (1982). Life span of the biosphere. *Nature, 296,* 561-563.

MacDonald, B. & Stronach, Y. (1989). The independent policy evaluation. *Journal of Computer Assisted Learning, 5,* 51-64.

Muñoz, E. (1998). Biotecnología: ciencia-técnica entre dos paradigmas. At *Modern and postmodern science.* Madrid: Fundacion March.

Prigogine, I. (1993). *Le leggi del caos.* Laterza.

Schröedinger, E. (1967). *What is life?* Cambridge Press.

Sáez M. J. y Carretero, A. (1995). Un modelo posible: la Evaluación en la España actual. En Sáez, M. J. (Ed.), *Conceptualizando la evaluación en España* (pp. 131-142). ICE Univ. Alcalá de Henares.

Scriven, M. (1997). Truth and objectivity in evaluation. In E. Chelinsky & W. Shadish (Eds.), *Evaluation for 21st century.* Thousand Oaks, CA: Sage.

Simons, H. (1987). *Getting to know schools in a democracy: The politics and process of evaluation.* London: The Falmer Press.

Snow, C. P. (1969). *The two cultures and a second look.* Cambridge University Press.

Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stake, R. E. & Easley, J. A. (1979). *Case studies in science education*. Urbana, IL: CIRCE.

Verdansky, V. (1926). The biosphere. In L. Margulis & D. Sagan, (1995) *What is the life?* New York: Nevraumont Publishing Company.

Wells, A. S., Hirshberg, D., Lipton, M., & Oakes, J. (1995). Bounding the case within its context: A constructivist approach to studying detracking reform. *Educational Researcher*, 24(5), 18-24.

Excerpts from An Evaluation of Kenwood Elementary School's Year-Round Program

Delwyn L. Harnish,[1] Philip P. Zodhiates,
and Najmuddin Shaik
University of Illinois

Background

The traditional nine-month school calendar long ago lost its reason for being. Originally designed to serve the cyclical manpower needs of the predominantly agrarian economy of the 19th century, the nine-month school calendar has remained unchanged, in large part, because of social convention rather than any economic or pedagogical imperative. Although the vast majority of schools continue to operate on a nine-month calendar, in recent years there has been an exponential growth in the number of schools that have adopted a year-round schedule.

Figure 1. Growth of Year-Round Education in the U.S.

School Year	States	Districts	Schools	Students
1985-1986	16	63	410	354087
1995	37	436	2252	1649380

Source: National Association for Year-Round Education

This movement away from the traditional school calendar and toward year-round schools is propelled by a number of factors, perhaps primary among them the concern with the learning loss that occurs among students during the long hiatus of the summer months. This learning loss is especially worrisome among pupils who are already behind their peers, and who each year fall further and further behind.

In an effort to address these concerns, in the summer of 1995 the Champaign School District established a year-round program at Kenwood Elementary School. Since then, a number of other districts in the area have followed suit; others are planning to do so in the near future. In the fall of 1996 we were asked to examine systematically the operation of this still new, experimental program.

---

[1] University of Illinois at Urbana-Champaign, College of Education, department of Educational Psychology, 1310 South Sixth Street, Champaign, IL 61820. (217)333-4416, harnisch@uiuc.edu

Methodology

Issues. We examined the impact of the new school calendar on Kenwood Elementary School in collaboration with Principal Les Huddle, the staff of Kenwood Elementary School, and district administrators. The study has examined the effect of Kenwood's year-round calendar in nine areas:

- student achievement
- student behavior
- students' families
- the curriculum
- teachers' perceptions of student learning and behavior
- teachers' sense of community
- teacher job satisfaction
- the role of administrators
- the school budget

Data Collection. The study made use of both quantitative and qualitative data in examining the effects of a year-round calendar in these nine areas. We designed and distributed surveys to teachers as well as to families of students. We made use of data gathered by the school on student achievement and student behavior. In addition, we conducted focus groups and individual interviews with students, teachers, and administrations.

Study Questions. The study examined the impact of the year-round calendar on student achievement and behavior, families, the curriculum, teachers, administrators, and school budget by focusing on the following list of questions:

- What changes, if any, have occurred in the wake of the new year-round program in student academic performance in the three main subject areas--mathematics, reading, and language arts?
- What impact has the new program had on student attitude and behavior?

- What are the benefits and costs of the new calendar for the families of the students?
- In the view of teachers, what difference has the new calendar made on student learning and student behavior?
- What impact has the new calendar had on the curriculum?
- What effect has the new calendar had on the professional development of the teachers?
- What difference has the new calendar made in teachers' sense of community?
- What are the benefits and costs of the year-round calendar for teachers?
- What additional responsibilities have administrators faced in moving to a year-round curriculum?
- What impact has the year-round calendar had on the school budget?

Summary of Findings

Teacher Comments

Teacher perceptions of student learning. Teachers reported that students are more likely to engage in continuous learning and that behaviors conducive to student learning tend to persist during intersessions and summer breaks. Teachers have no clear sense, however, of whether the year-round calendar has made a difference in students' academic performance. The most frequent comment teachers made regarding the impact on student learning of the YR calendar is that students seem to forget less of what they have learned from one year to the next. Regular classroom teachers reported smaller learning losses among students after intersession and summer breaks. A number of teachers also noticed a positive change in students' energy level.

Despite their enthusiasm for the YR program, none of the teachers interviewed were able to say with any certainty that the new calendar has resulted in increased student learning. As one teacher put it, "The whole reason we did this--we told parents--is so their kids will learn better. I'm hard pressed to find people working here who want to go back to the traditional calendar. Kids really enjoy it. But I'd like to show it makes a difference in kids' learning, in their academic performance. I'd like to see [test scores of] kids now in second grade, who started year-round in kindergarten. If we can show in test scores that kids learn more, that's the bottom line."

Teacher perceptions of student behavior. Although student behavior problems have not gone away, they are less likely to escalate out of control as a result of the nine weeks on and three weeks off YR schedule. Teachers seemed to agree that the YR calendar has had some effect on student behavior, although they have difficulty sorting out the impact of classroom dynamics from those of the year-round calendar itself. Nonetheless, a number of teachers suggested that the pattern of nine weeks on/three weeks off is a factor in student behavior.

Teacher perceptions of the impact of YR on curriculum. Teachers feel that the calendar encourages them to make better use of instructional time, to organize their curriculum more effectively, to reflect on what they are doing, and to make changes when necessary. They seem to value especially the opportunities offered by the calendar for on-going planning and self-evaluation throughout the year. However, our analysis indicates that the opportunities offered by the new calendar have not been fully realized for curriculum planning and teacher self-evaluation.

A number of teachers pointed out that the schedule of nine weeks on/three weeks off gives them a chance to evaluate how things are going in their classroom and to revise their curriculum to fit their students' needs. Teachers also credit the new calendar with encouraging them to make more effective use of their instructional time. While other schools by and large stop teaching new material in early May, Kenwood teachers have told us repeatedly that they and their students continue to do serious work up until the end of the school year.

Teacher perceptions of the impact of YR on professional development. Teacher comments regarding professional development fall into two categories: graduate courses offered by the University of Illinois at Urbana Champaign (UIUC), and in-service programs organized by the school district. In both cases, teachers see advantages and disadvantages in the YR calendar.

For teachers who are pursuing a higher degree, or who are enrolled in continuing education courses, Kenwood's calendar has some advantages and some disadvantages. One benefit of three-week intersessions is that they provide ample time to complete class assignments. Even teachers who are not enrolled in a university course report that they do more professional reading than they did under the traditional calendar. Kenwood's short summer break, however, makes it difficult for Kenwood teachers to take summer courses offered by UIUC. These summer courses, scheduled with the traditional school calendar in mind, do not always line up well with Kenwood's schedule.

A number of teachers who are either currently enrolled at the UIUC, or who have taken courses in the past, commented on the positive impact of the YR calendar on their lives. One teacher reported that it is now easier to take regular university courses offered during the academic year because she now has a three-week block of time when she is free of teaching responsibilities and can devote herself to her academic work. Other teachers have also found ways to fit their university courses into their teaching schedules.

Other teachers, however, voiced concerns about juggling the demands of work and school. The issue seems to be the fit between Kenwood's six-week summer break and the scheduling of required summer courses at UIUC. Kenwood teachers seem to be at a disadvantage compared to other teachers when it comes to summer courses. What happens, some teachers asked, when one is not in a position to pick a four-week summer course that falls within Kenwood's six-week break?

The school district offers a variety of professional development programs for teachers, such as short courses in computer literacy. Kenwood teachers say that one aspect of their schedule in particular makes it easier for them to participate in the district's in-service classes: Kenwood's classes, due to Champaign's school bus schedule, begin at 8:10 a.m. and end at 2:00 p.m. Many other teachers have praised this feature of the Kenwood calendar, especially because they have the rest of the day to themselves.

We also heard a number of complaints about the fit between Kenwood's YR calendar and the school district's schedule of in-service classes. The problem seems to occur when a district in-service course is offered during one of Kenwood's intersessions, which Kenwood teachers view as their time off. There seems to be a widespread feeling among Kenwood teachers that the school district forgets them when it comes to scheduling in-service learning opportunities or meetings for teachers. When in-service classes occur during Kenwood's intersessions, teachers are faced with a dilemma: Do they sacrifice their time off for the sake of professional development, or do they give up a learning opportunity in order to preserve their free time?

This on-going conflict between Kenwood's calendar and the district's schedule of meetings and classes cause Kenwood teachers to have mixed feelings about their sense of professionalism. It is not surprising, therefore, that this issue is troubling to Kenwood teachers.

Teacher perceptions of the impact of YR on their sense of community. Kenwood's status as the only YR school within the school district contributes both to a sense of pride and to feelings of invisibility and isolation among Kenwood teachers. Problems resulting from the district's scheduling of professional development activities have caused resentment among some Kenwood teachers. A feeling that they are invisible seemed to color a number of teacher comments. In part, the sense of isolation seems to come from what teachers perceive as the public's misapprehensions about YRE.

Although some teachers complained about being overlooked or ignored by the rest of the district, most Kenwood teachers seem to take great pride in being special. Teachers feel a powerful sense of ownership and pride that was heard in many teacher comments.

Impact of YR on teachers' lives: Benefits and costs. Perhaps the strongest finding of our with teachers, is that they report a more positive attitude toward their work as a result of the YR calendar. They suggest that it has contributed to less stress and less teacher burn-out. The YR calendar enables teachers to sustain a high level of energy in their work, and a healthy balance between work and home.

One of the major consequences of the YR calendar, according to all the teachers we interviewed, is less teacher burn-out. Reduced job stress is another major theme. Teachers reported that the YR calendar, in contrast to the traditional calendar, allows them to have a life outside of school. This greater sense of balance between school and home contributes to a more positive attitude toward their teaching, and to a higher energy level at work. Other teachers spoke of having more time to devote to house projects such as gardening or sewing, something they couldn't do before.

Teacher perceptions of intersession. Intersession originally was conceived as a kind of summer school for students who needed remediation, and for other students, it was seen as an opportunity for enrichment. According to a number of teachers, its potential has not been fully realized. The main obstacle appears to be adequate staffing. Teachers described how the school always seems to struggle to find teaching staff for the various programs.

Teachers spoke of their wish to use intersession more effectively for at risk students. A solution, according to some teachers, would be a smaller ratio of students to teachers. The current size of Recovery classes is typically between 17 to 20 students, requiring the teacher to do little more than maintain control over the students during the one and a half hour sessions. Suggestions teachers offered to improve the Recovery Program include providing more tutoring or small-group instruction, offering transportation for families who need it, and increasing the amount of class time per day.

Parent Comments

Neighborhood school or school of choice? After three years as a school of choice with a YR calendar, Kenwood continues to be perceived primarily as a neighborhood school. In 1995-1996, the first year of the year-round calendar, 75 percent of the students attending Kenwood lived in the neighborhood. Three years later, in 1997-1998, that percentage had dropped slightly to 70 percent. Even the parents who live in the Kenwood catchment area, however, appear to be enthusiastic about the YR calendar, and Kenwood's identity as a YR school.

Parent perceptions of the impact of YRE on regular students. Kenwood parents reported that their children seem to retain more of what they have learned, and spend less time reviewing old material. Most parents, however, were not able to compare the impact on learning of YRE to the traditional calendar. Although 90 percent of the parents surveyed saw YRE as an effective educational program, we found it difficult to sort out the facts from the rhetoric and assumptions surrounding YRE.

One parent told us that she likes the year-round schedule because students retain more and spend less time on reviewing, devote more time learning. She said, "the teacher has more time to cover more material in the class." Another parent said that students are less burnt-out and they can relax during the intersession breaks.

Parent perceptions about the impact of YRE on special needs students. A number of parents reported that the YR program at Kenwood has benefited their special needs. They talked about seeing positive changes, such as improved motivation and performance. However, the factors to which they attribute these improvements--working closely with a teacher; a teacher helping diagnose their child's condition; putting their child on medication--are not specific to YRE. We also found that a small number of parents expressed some anxiety and frustration about the inability of some of the teachers to work with their special needs children. Here, too, parent concerns seemed to have little to do with YRE.

Parent perceptions of the impact of YRE on student behavior. Most parents perceived no impact of YRE on their children's behavior. Although 46 percent of parents surveyed reported an improvement in their children's school behavior, when they talked about their own child's behavior, or about their concerns with the behavior of other children, they did not make a positive or negative connection with YRE.

The impact of YRE on parents' and students' attitudes toward school. Most parents reported that their children have positive feelings toward school, and that they look forward to going to school. Of the parents surveyed, 84 percent reported that their children looked forward to going to school every day. Only a few parents, however, make the connection between their children's positive attitude and YRE.

Parents as well as children seem to have a more positive attitude toward school. This positive attitude, Les Huddle believes, is due in large part to the fact that Kenwood is a school of choice. Parents appear to take more ownership of the school, which is evident from the higher attendance at PTA meetings, and an increase in the number of parent volunteers in the classroom. Parents are more willing to identify with the school mission and provide support in many different ways.

Intersession. Although most of those parents who enroll their children in intersession programs are pleased with the experience, the majority of parents we talked to have not made use of the intersession programs that are available at Kenwood. Parents who have not participated in intersession classes cite three barriers: (1) tuition costs, (2) lack of transportation, and (3) scheduling.

The impact of YRE on the families of students. Parents with children in both a YR and a traditional school program reported little inconvenience, and indicated that they had experienced no problems planning family activities, such as vacations. Of the families surveyed, 79 percent applauded Kenwood's efforts to directly involve them in their children's education, and 81 percent reported that they enjoyed their involvement with the school's YR program.

Parent comments about the administration and teachers. The parents had very positive things to say about the principal of Kenwood. Parents frequently characterized the principal as approachable, responsive, and caring. The great majority of parents have positive things to say about the teachers at Kenwood. They noted that teachers help diagnose ADD, work with parents to improve students' academic performance, and keep lines of communication open with parents. Of the parents surveyed, 81 percent reported their belief that the Kenwood staff really cares about the welfare of their children. A few parents of children with ADD, however, are unhappy with the way some teachers deal with them and with their children.

Student Comments

Student attitudes toward YRE. In part, students' enthusiasm for the new calendar, and the sense that it made them special, seemed to be based on what one might call "creative mathematics" about the amount of vacation time Kenwood students have. Objectively, of course, Kenwood students attend school for 180 days, exactly the same number as all other students in Champaign. Nonetheless, many of the students we talked to had a different perception of the arithmetic of school attendance at Kenwood. "You get more time off due to the 9-week classes and 3-week break," Stan said. "We can tease other friends that they are still going to school when we have break." Kevin agreed: "It's cool because you get more time off for vacation." Then he added, "Instead of one long vacation time we have two short vacations. It doubles our vacation time."

This common, though mistaken, perception of the benefits of Kenwood's schedule seems to be part of an amorphous agglomeration of beliefs and understandings about the advantages of YR education. For the students, these perceived benefits center around increased vacation time; for teachers and parents they revolve around improved learning and greater curriculum coverage. These beliefs are what we have come to call "the ideology of YR education." Like any other kind of ideology, it binds people together in a community of shared beliefs and understandings, and contributes to a group's sense of common purpose and high morale. We found it noteworthy that both students and adults at Kenwood shared a common set of beliefs and attitudes about the benefits of the YR calendar. This shared ideology, we believe, both nurtures and is sustained by the students' and adults' feeling of uniqueness within the Champaign school district.

Student comments not specific to YRE. As one might expect, students had difficulty making distinctions between their impressions of Kenwood School and their evaluation of the year-round calendar more specifically. For example, some students talked about their preference for Kenwood's early daily schedule, which starts at 7:45 AM and ends at 2:05 PM; students liked going home forty minutes earlier than other schools. They talked about having more time to "do fun things" and "more time to do the homework." Although an important feature of Kenwood, early dismissal has little to do with the year-round calendar itself. (To some extent, this inability to distinguish those aspects of Kenwood that are particular to the year-round calendar from other features of the school was also apparent in the comments teachers and parents made about Kenwood.)

Carmen Palmer

Mary Ann Ludwig

Lois Johnson Gueno

A Study of an Empowered School: An Investigation of the Development and the Effect of a Teacher Empowerment Process

Carmen L. C. Palmer
University of Illinois

The purpose of the study was to describe the evolution of the school's advisory teacher empowerment process, to present an illustrative example of its use, and to investigate the effect of its use to propose school policy designed to solve teacher identified problems and concerns perceived to negatively affect their professional performance and effectiveness.

School reform proposals accepted the theory that schools work best when teachers and principals are more involved in the problem solving at the building level. Proposals called for "increased participation in decision making" (Phillips, 1989, p. 3). Chicago Public schools were mandated into school reform by the Illinois General Assembly in 1988 with the passing of P.A. 85-1418. It included a provision for an advisory teacher empowerment component.

The professional personnel advisory committee, the PPAC, is the advisory teacher empowerment avenue provided by state legislation. Every school is required to elect a PPAC each year for the purpose of advising the principal and the mandated popularly elected local school council (LSC), the school's governing body.

This paper addressed two main questions:

1. What does a faculty of an elementary, urban magnet school do when provided an opportunity for legislated teacher empowerment?

2. What model of a teacher empowerment process emerged?

Over a span of 5 years, and with the legislative mandate, the faculty of the school advised the principal and the LSC on several teacher identified school-based reforms.

Figure 1. Summary of a five year historical review of what the empowered school did with its legislated advisory teacher empowerment opportunity.

Year 1		
	Events Initiating Empowerment	Principal Retirement
		Teacher Catalyst Response to School Reform
	Activities	Organized meeting schedules, places, times.
		Surveyed for principal qualities preferred.
		Partnered with LSC on principal selection process.
	Outcome	Increased teacher participation in principal selection committee
Year 2		
	Focus	Principal selection
	Activities	Continued to organize; identified PPAC roles and responsibilities; established PPAC committees
	Outcome	An 18-step principal selection process, a principal selection that reflected the teachers' preferences
Year 3		
	Focus	School security
	Activities	Continued to organize; developed data-collection, analysis, and intervention activities
	Outcome	Findings: in-school security = very good; out of building = 3 areas to address; identification of monetary cost-free solutions.
Year 4		
	Focus	Faculty concerns
	Activities	Improved data-collection, analysis and intervention activities
	Outcome	Identification of nine categories of concern; emergence of new leaders in the form of 5 new committees and volunteer chairmen.
Year 5		
	Focus	Interruption of AIT due to SUI
	Activities	Began to negotiate implementation of TPSP
	Outcome	TPSP implemented

These are outlined in figure 1. In year four, after feeling very encouraged by the empowerment successes of the previous years (including participation in designing an 18-step principal selection process, and resolving safety needs of the school) a survey of faculty concerns was developed by the faculty to identify the next "teacher empowerment" focus. Nine categories of concerns were identified. Interruptions to instructional time was the most salient teacher concern.

The interruptions reported were ranked in order of importance by the faculty. Of the reported interruptions, intercom interruptions were the highest ranked instructional interruption, and therefore became the teachers' empowerment focus for year five.

In order to better define the problem of interruption to allotted instructional time due to the school's use of the intercom, the instrument "Intercom Interruption Tally Sheet" was created. With the aid of 12 faculty members, one from each of the 8 grade levels and 4 from educational resource programs, data defining the schools use of the intercom was collected and analyzed. The most important findings were that 96% of the 61 intercom interruptions for the week were made by the administration with an average frequency of 12 interruptions per day, 3 occurring on average during the first period which was the schoolwide reading period.

The teacher proposed school policy advisements (all of which were implemented by the school's administration) were:

1. Make better use of the non-instructional time BEFORE first period reading for intercom announcements (7:45-8:00 a.m.).
2. Provide walkie-talkies for most paged staff.
3. Identify a process and standing time period that is established for school-wide announcements.
4. Limit request for reports to specific classrooms rather than school-wide announcements.

The Model of a Teacher Empowerment Process that Emerged. Figure 2 presents The Palmer Model of Teacher Empowerment. Figure 3 diagrams how the model is used. The data-driven teacher empowerment process that evolved in the school successfully facilitated the teacher empowerment activities by producing teacher proposed school policy. The policy was perceived by the faculty to solve data-supported, teacher-identified problems. The Palmer Model explains the stages of creating the new teacher proposed school policy--from creating the advisory packets and having them approved by the faculty, and reviewed by the principal to finally being presented to the school's governing body. When this model was statistically tested with a t-test, the model was found to be effective at a 0.1 level of significance when addressing interruptions to allotted instructional time due to the school's use of the intercom.

Figure 2. The Palmer Model of Teacher Empowerment

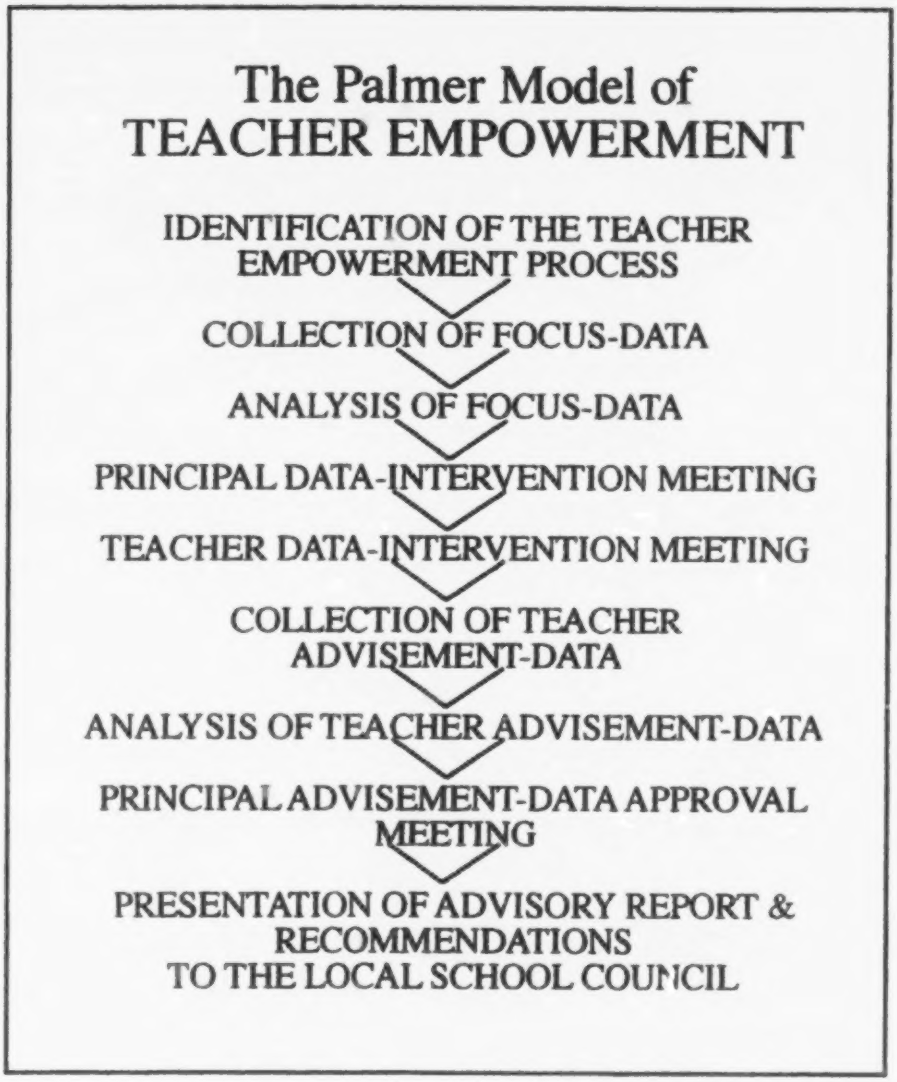
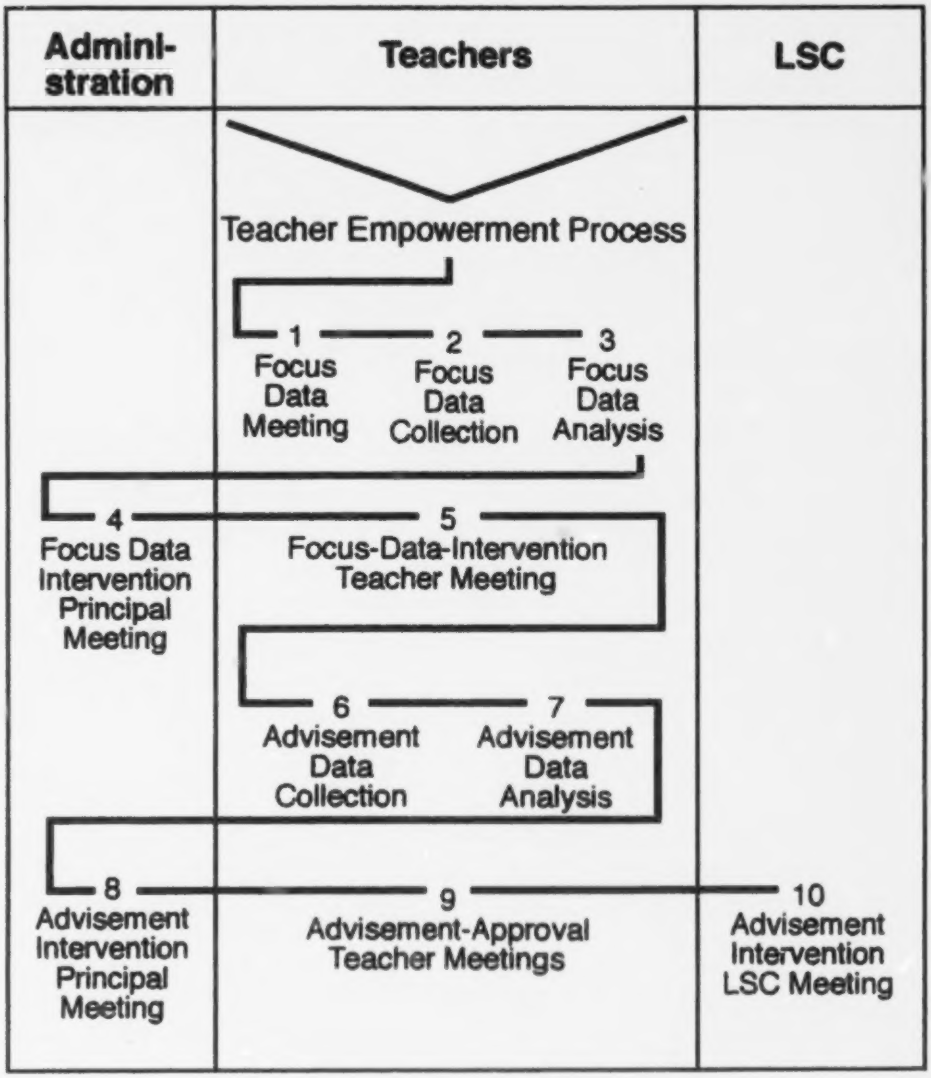

The Palmer Model of
TEACHER EMPOWERMENT
IDENTIFICATION OF THE TEACHER EMPOWERMENT PROCESS
COLLECTION OF FOCUS-DATA
ANALYSIS OF FOCUS-DATA
PRINCIPAL DATA-INTERVENTION MEETING
TEACHER DATA-INTERVENTION MEETING
COLLECTION OF TEACHER ADVISEMENT-DATA
ANALYSIS OF TEACHER ADVISEMENT-DATA
PRINCIPAL ADVISEMENT-DATA APPROVAL MEETING
PRESENTATION OF ADVISORY REPORT & RECOMMENDATIONS TO THE LOCAL SCHOOL COUNCIL

Figure 3. A diagram of the use of the Palmer Model of Teacher Empowerment


Administration
Teachers
LSC
Teacher Empowerment Process
1 Focus Data Meeting
2 Focus Data Collection
3 Focus Data Analysis
4 Focus Data Intervention Principal Meeting
5 Focus-Data-Intervention Teacher Meeting
6 Advisement Data Collection
7 Advisement Data Analysis
8 Advisement Intervention Principal Meeting
9 Advisement-Approval Teacher Meetings
10 Advisement Intervention LSC Meeting

Conclusions

In the empowered school, when the faculty was given the teacher empowerment opportunity of shared decision making school based management, teachers demonstrated that a teacher empowerment process can be employed to solve problems. The study also provides indirect indications that teacher empowerment, practiced as described in the study's model, can make the environment of the school more conducive to teaching and learning. These teachers positively influenced: (a) the principal selection process; (b) school safety; (c) how to determine teacher concerns and, (d) how to reduce interruptions to instructional time.

School reform is demanding that future teachers be prepared to be educational leaders and to participate in policy making activities. It will not matter how well prepared our pre-service teachers are in their content area and methodologies, if they are ill-equipped to maneuver through the political realities of the school houses and districts in which they teach.

Future teachers will not be able to enjoy the experience of teaching to their professional best if they are ignorant of the policies that govern their profession. They must be well aware of their teacher empowerment tools such as their contracts, school district policy, and school law. When teachers are placed in frameworks which prevent them from teaching to their professional best, *children suffer academically.* Therefore, teachers must be aware of the policy that sets the political framework in which they teach and in which they will have to address educational issues as they evolve in their schools.

As a veteran educator in the public schools for some 28 years, I charge the institutions of higher education with the task of producing the educators that school reform is demanding--aggressive educational leaders who not only love children *but who are grounded in their content area as well as in educational policy and who understand that they will have to fight for children by fighting for the professionalization of teaching via policy making activities.*

Effects of a Museum-School Collaborative on Seventh Grade Students of an Urban Public Elementary School

Mary Ann Ludwig
Chicago Public School

I'm sure at some time in your life, probably many times, you've found yourself standing in front of an art object in a museum totally engrossed in what you're viewing-- "bonded to it but not in bondage," says Crowther, 1993, in *Art and Embodiment: From Aesthetics to Self-Consciousness.*

> Rather, the contents of the present are opened up as a zone of pure explorative possibility in perceptual terms. We are active, we have an enhanced sense of life precisely because the conditions and burdens objectively placed on the exercise of freedom are lifted. We experience freedom in an enhanced form (Crowther, 1993, p. 160).

In this case study, I hypothesized that direct experience with objects in an art museum that link history with art, the past with the present, and art with other aspects of life, can contribute to students' ability to experience a more realistic, personal, and integrated understanding of life and times in the past and present.

Student thinking might be expanded so that they no longer see subject areas as separate categories, but as part of a larger whole because boundaries have been removed. This was an exploratory study to determine if there was some indication that the vitalizing effects suggested above could be seen in a particular museum-school collaborative involving only a single museum visit.

The possibility of connecting learning to personal experiences was the aim of a Museum Classroom project called "American Art and Culture: 1650-1993," which attempted to integrate museum methodologies and materials into the standard 7th and 11th grade curriculum. The collaborative was funded by a grant from the National Endowment for the Arts. It drew on The Art Institute of Chicago's strong permanent collections of American art and made use of the newly renovated Kraft General Foods Education Center. Six Chicago Area schools, three suburban high schools and three Chicago elementary schools, participated with different curricular emphases that were embodied in interdisciplinary curricular plans. Specifically, my study focused on the experiences of seventh graders of an urban public elementary school whose curriculum focus at the time was Colonial America.

The samples of student work were analyzed (a) sketchbook/journal notes, (b) creative and descriptive writing about a portrait of "Mary Greene Hubbard" by John Singleton Copley, and (c) art work produced. In addition, interviews with some students were conducted and analyzed.

Analysis of Student Art Products for Evidence of Connection with Art Institute Experiences[1]

The effects of Museum Classroom, "American Art and Culture: 1650-1994" on seventh grade students of Beasley Academic Center are evident in the art they made in connection with the interdisciplinary experience at the Art Institute of Chicago. Each of the five groups focused on a different medium in their preparation, tour, studio experience, and follow-up activities. Following is a discussion of examples from each group.

---

[1] The presentation at the Stake Symposium included 23 samples of student artwork. For purposes of space in this collection, 5 samples are included, one sample for each art medium studied by the researcher.

Group 1 - Furniture
Most students from this group constructed chairs, chests, cabinets, beds, dressers, or wardrobes from cardboard using paint, markers, and fabric to add details.

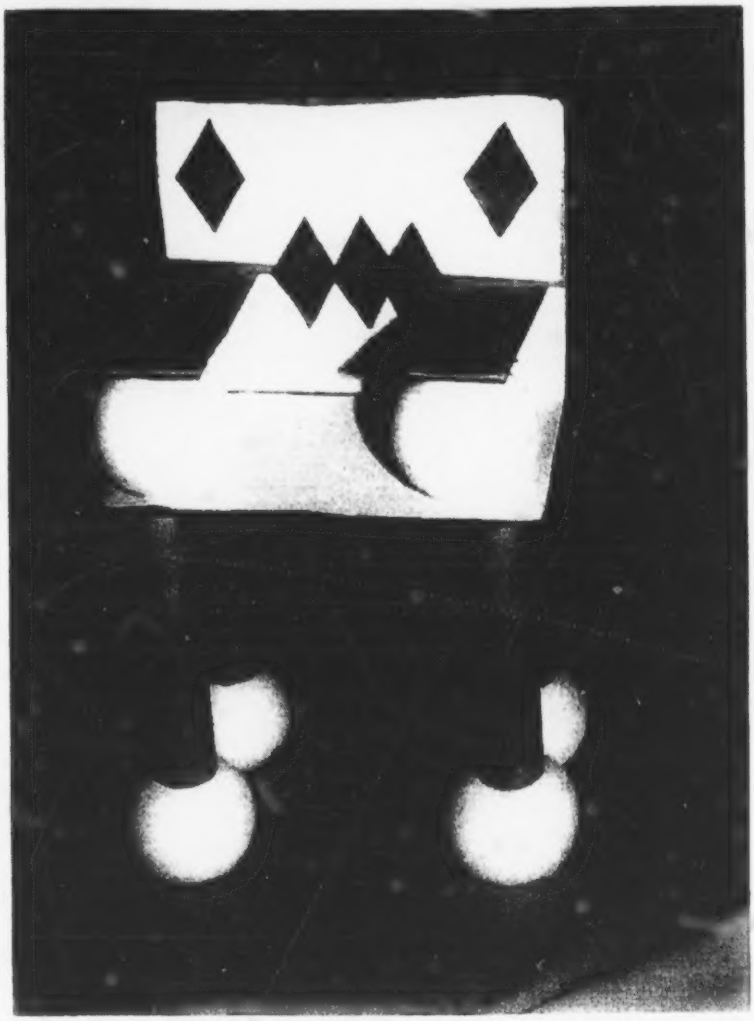


Figure 1. White straight-backed arm-chair with ball-and-claw feet.

The piece shows that the student constructing it synthesized information acquired through doing the worksheet about the chair on the bus and touring the early American galleries and Thorne Rooms. The ball-and-claw were common features on Chippendale furniture of early America that students saw and which was explained by the docent. The red diamond shaped designs were a type common to chairs of the times. They were copied by their creators from pattern books that were circulated through the New England, Middle, and Southern colonies.

Group 2 - Architecture
Students in this group used pre-drawn white, cardboard shapes to cut and fold into houses for a neighborhood. They used tempera paints, watercolors, and markers to apply details after the doors and windows had been drawn on the sides of the building.

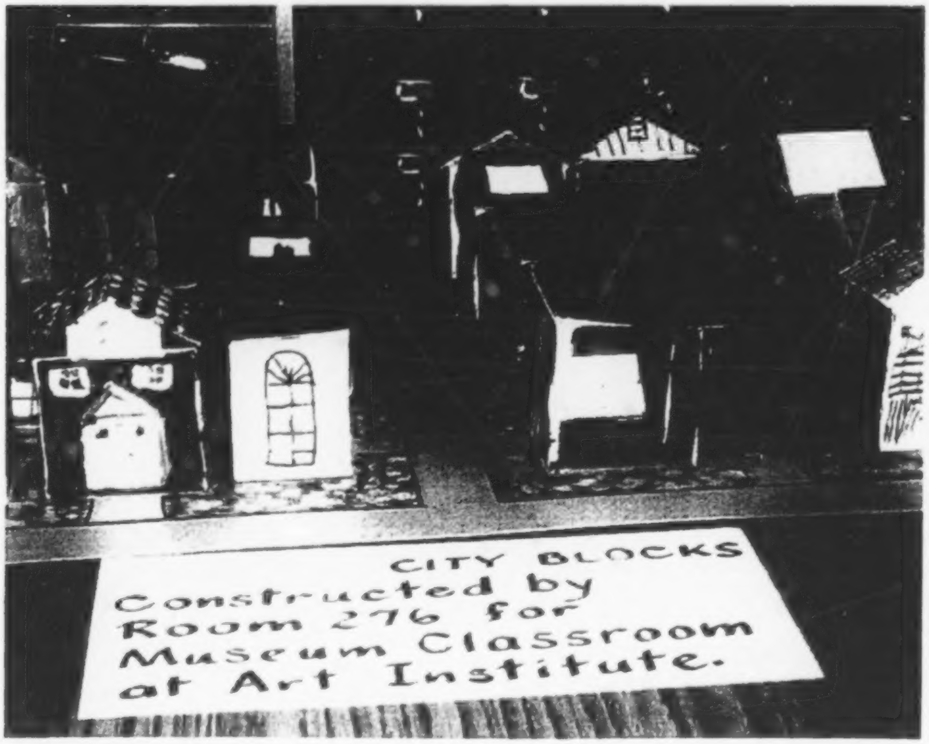


Figure 2. Students' neighborhood.

Buildings in this city block show evidence of the architectural elements students saw and discussed with the docent in the galleries and with the teacher in the studio class: columns, pediment, arch, small-paned windows, double doors. Their own homes don't look like this, but when given the chance to construct their own building for the first time, the seventh graders employed the new concepts they had learned.

Group 3 - Landscape

Students from this homeroom listened to a passage by Thoreau read by the teacher, and interpreted it in pastel compositions.

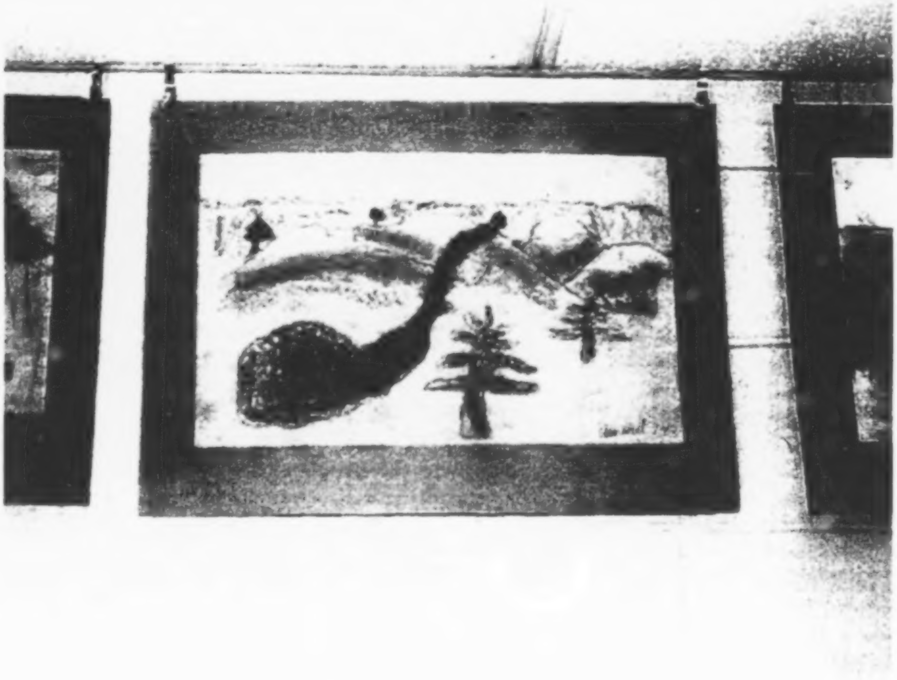


Figure 3. Stream through mountains.

This art work shows a stream coming toward the viewer through the pastel landscape. The student has placed a tree in the foreground, mountains in the middle ground, and sky in the background making use of the concepts stressed by the docent on the tour of the early American galleries. Distance is also shown by having the stream narrow as it recedes into the background uniting the parts of the composition.

Group 4 - Portraiture
This group of students did mod portraits with Polaroid snapshots at the Art Institute, and magazine self-portrait collages at school.

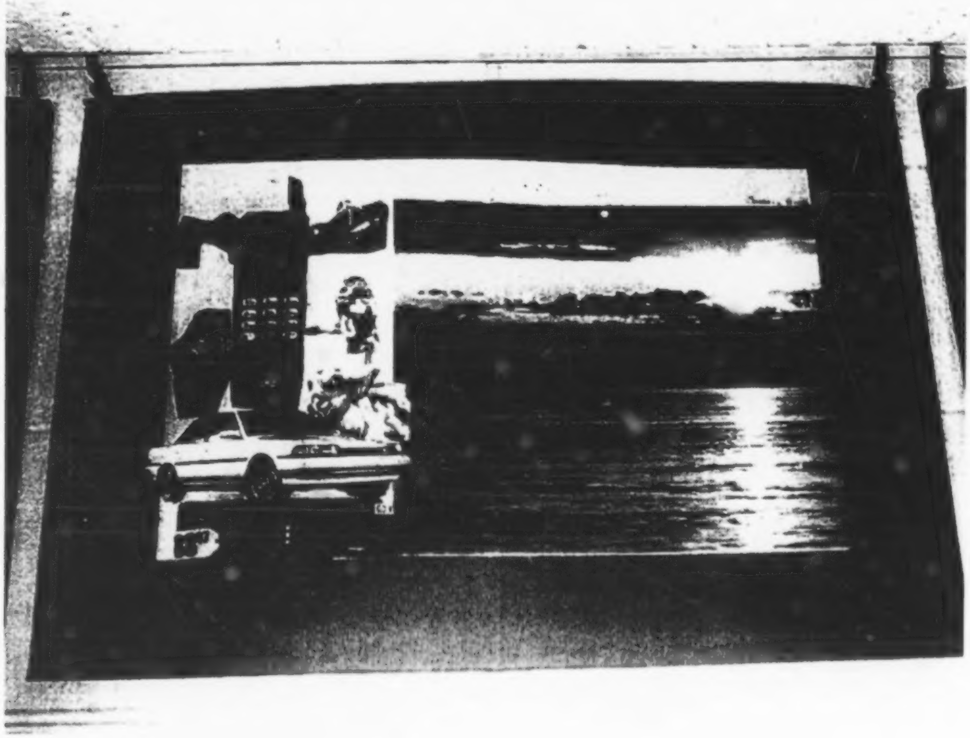


Figure 4. My favorite things collage.

This young man has a diversity of interests from hot cars to a cool sunset over the lake. He has used informal balance by placing a lot of smaller things compactly on the left side to balance the large expanse of water that makes up the rest of the composition.

Group 5 - Printmaking
This group did texture rubbings at the Art Institute and cardboard prints in art class at school.

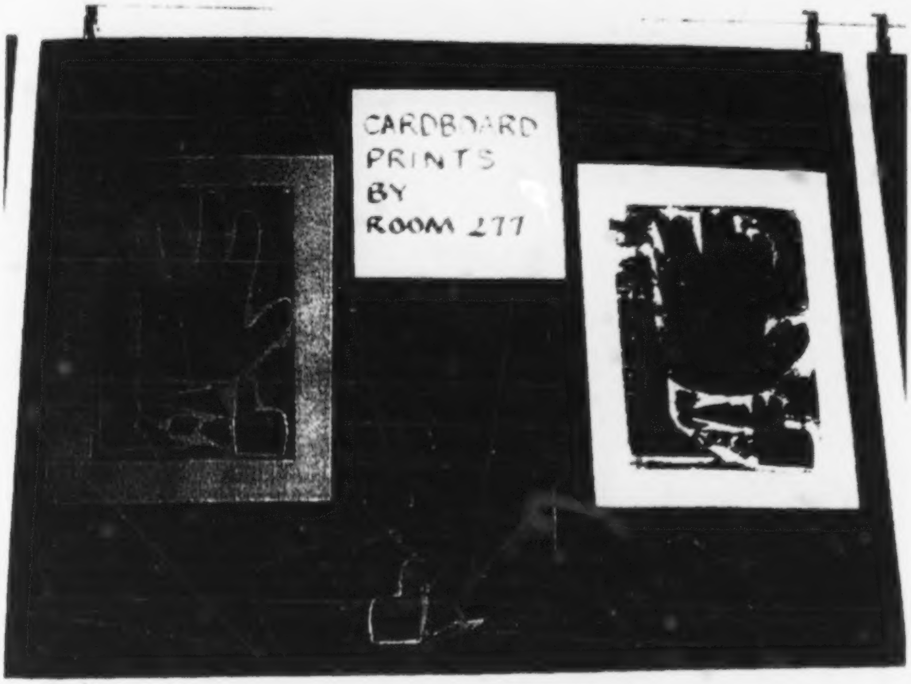


Figure 5. Lend me a hand

In this composition the art student uses the basic shape of a hand to experiment with the process of printmaking. One color and two color prints are displayed along with the original plate.

Conclusions

This study of the effects of a museum-school collaborative shows that meaningful learning takes place when the cognitive, affective, and motor-skill faculties are engaged in an interdisciplinary experience. Preparation is important by all involved, especially the museum education staff and the school teachers. When themes are carefully selected, goals and objectives made clear, authentic objects presented in context with information about the culture involved, students are able to have meaningful aesthetic as well as intellectual experience, the results of which are both immediate and lasting. With opportunities for reflection, discussion, and creative response in the form of writing and art making, meanings are synthesized and larger understandings attained.

Two Faces of Urban High School Students: Characteristics of Dropouts and Persisters

Lois M. Johnson Gueno
Chicago Public Schools

This study grew out of a curiosity I had about students who attended my school. I was a counselor at a school where approximately one half of the students who entered dropped out before graduation. I was curious as to why some stayed and some did not. The school community was not unusual, not unlike many urban school communities in the same region of the city. The school curriculum was a general one taken by all students and similar to that of most general high schools. The school was located on the south side of Chicago.

Given the profile of this community, the following research questions evolved to guide the study:

1. What are the characteristics of the 1989 cohort in terms of available school records?
2. How does a researcher access dropouts and persisters in an urban setting?
3. What are the characteristics and perceptions of the students who dropped out of this urban high school from the 1989 cohort?
4. What are the characteristics and perceptions of the students who persisted at this urban high school from the 1989 cohort?
5. How do the characteristics and perceptions of the dropouts and persisters compare?

The entire cohort consisted of 301 first-time 9th graders. For the purpose of the study, the cohort was grouped into several categories:

Table 1. Descriptors of the Cohort

301 entire cohort	174 males	127 females (42.2%)
54 (18%) still enrolled*	33 males (61%)	21 females (38.9%)
104 (35%) transferred out	63 males (60.6%)	41 females (39.4%)
70 (23% persisters (grads)	29 males (41.4%)	41 females (58.6%)
73 (24%) dropouts*	49 males (67.1%)	24 females (32.9%)

* Still enrolled = continued enrollment after graduation;
* Dropouts = those who left before graduation.

The characteristics of gender, entry age to 9th grade, entry reading stanine, and elementary school mobility were identified. Cross-classifications were done on characteristics found associated with dropouts--gender with overage, age with stanine, mobility, and dropouts.

I began my search for cohorts by contacting the Research, Evaluation, and Planning Department of the Chicago Public Schools. They provided me with printouts of the cohort identifying addresses and phone numbers. Locating dropouts and persisters proved to be extremely difficult. Four dropouts and the same number of persisters were found and studied.

I found that the parents of both groups (dropouts and persisters) expressed a desire for their children to receive a good education and gave what support they could. When the profiles of the two groups were compared, however, there were some differences, and they are presented in the chart below:

Table 2. Profiles of Dropouts and Persisters

	Dropouts	**Persisters**
gender	male	male
age	15 yr. of age (14.75)	14 years of age (14.25)
retentions	1.5 grades in elementary school *	no significant grade retention (.25) *
mobility	3.5 schools prior to HS *	1.25 schools prior to HS *
stanine-reading-9th	3 *	4 *
GPA	.434 (F) *	2.195 (C) *
absence final semester	44.6 days *	4.25 days *

income	from a low income family	1/2 low income
exit age	dropped out at 10th grade at age 17 (17.25)	graduated at age 18
Kindergarten (CPS)	equally divided between those who attended kindergarten in the Chicago Public schools and those who did not	attended kindergarten in Chicago Public Schools
Parent Educational Attainment	Half of the dropouts unaware of educational attainment of parents	Majority graduated high school and had some post high school education.
Frequent reasons for dropping out	fatherhood, motherhood, or employment were not reasons for dropping out	fatherhood, motherhood, or employment were not reasons for persisting in school

* = Average

Findings on the perceptions of the students on a number of variables showed there was no expressed difference of safety in and around school in terms of gangs and drugs preventing school attendance in either group. Their expressed feelings of safety in general, however, differed. One of the four dropouts said he did not feel safe while three of the four persisters said they did not feel safe. Dropouts had experienced more incidences of suspension than the persisters. All four dropouts had been suspended in elementary school and three had been suspended in high school. Two of the persisters had been suspended at each level.

It was found that persisters perceived treatment by teachers and counselors to be more fair than the dropouts. The two groups were divided in their perception of principal fairness. Persisters perceived discipline to be less fair and more effective than dropouts. In terms of coursework, both groups considered fine arts and vocational classes to be important.

When asked about recommendations to reduce the dropout rate both groups suggested more counseling to talk about personal problems with counselors and teachers. Dropouts talked about cutting out gangs and persisters talked about a need for more police.

All of the dropouts and all of the persisters indicated that family members had helped with their schooling during their elementary school years. During the high school years, however, two of the dropouts reported receiving help, while three of the four persisters continued receiving help in high school. Both groups felt, equally, that for the most part, students drop out simply because they do not want to go to school. The persisters indicated that dropping out of school was never an issue. One half of the persisters attributed their graduating to a strict mother/parent(s) who cared about their graduating.

It is important to note when looking at the entire cohort, that the dropout rate for those at stanine four as well as those above stanine four was 42.4%. The lower the stanine the higher the dropout rate until stanine six and above. At stanine six and above, the dropout rate increased to 53.8% while students with stanine five were more likely to transfer out. It appears, therefore, that those who tested poorly dropped out at the highest rate, but the second highest dropout rate was among males who tested best. The best prepared students left, especially males. This school lost their students who scored highest as well as those who scored lowest.

In light of these findings, it is recommended that school counselors be alert to incoming students who possess characteristics typical of this dropout group. These characteristics may serve as markers for counselors to be alert to incoming students so that these students do not get further behind or are not adequately challenged academically and become potential dropouts. Counselors should be alert as well to the high scorers so they might be retained in the school.

Additionally, there needs to be a rethinking of the role of parents in the decisions of students to drop out of school. This study found that parents of both groups expressed strong interest and support for their children to stay in school and to perform well, yet a large number of students still dropped out.

And the more personal begin here

Gene Glass

David Balk

Tom Maguire

Ghosts and Reminiscences: My Last Day on Earth as a "Quantoid"

Gene V Glass
Arizona State University

I was taught early in my professional career that personal recollections were not proper stuff for academic discourse. The teacher was my graduate adviser Julian Stanley, and the occasion was the 1963 Annual Meeting of the American Educational Research Association. Walter Cook, of the University of Minnesota, had finished delivering his AERA presidential address. Cook had a few things to say about education, but he had used the opportunity to thank a number of personal friends for their contribution to his life and career, including Nate Gage and Nate's wife; he had spoken of family picnics with the Gages and other professional friends. Afterwards, Julian and Ellis Page and a few of us graduate students were huddled in a cocktail party listening to Julian's post mortem of the presidential remarks. He made it clear that such personal reminiscences on such an occasion were out of place, not to be indulged in. The lesson was clear, but I have been unable to desist from indulging my own predilection for personal memories in professional presentations. But that early lesson has not been forgotten. It remains as a tug on conscience from a hidden teacher, a twinge that says "You should not be doing this," whenever I transgress.

Bob Stake and I and Tom Green and Ralph Tyler (to name only four) come from a tiny quadrilateral no more than 30 miles on any side in Southeastern Nebraska, a fertile crescent (with a strong gradient trailing off to the northeast) that reaches from Adams to Bethany to South Lincoln to Crete, a mesopotamia between the Nemaha and the Blue Rivers that had no more than 100,000 population before WW II. I met Ralph Tyler only once or twice, and both times it was far from Nebraska. Tom Green and I have a relationship conducted entirely by email; we have never met face-to-face. But Bob Stake and I go back a long way.

On a warm autumn afternoon in 1960, I was walking across campus at the University of Nebraska headed for Love Library and, as it turned out, walking by chance into my own future. I bumped into Virgina Hubka, a young woman of 19 at the time, with whom I had grown up since the age of 10 or 11. We seldom saw each other on campus. She was an Education major, and I was studying math and German with prospects of becoming a foreign language teacher in a small town in Nebraska. I had been married for two years at that time and felt a chronic need of money that was being met by janitorial work. Ginny told me of a job for a computer programmer that had just been advertised in the Ed Psych Department where she worked part time as a typist. A new faculty member--just two years out of Princeton with a shiny new Ph.D. in Psychometrics--by the name of Bob Stake had received a government grant to do research.

I looked up Stake and found a young man scarcely ten years my senior with a remarkably athletic looking body for a professor. He was willing to hire a complete stranger as a computer programmer on his project, though the applicant admitted that he had never seen a computer (few had in those days). The project was a monte carlo simulation of sampling distributions of latent roots of the B* matrix in multidimensional scaling--which may shock latter-day admirers of Bob's qualitative contributions. Stake was then a confirmed "quantoid" (n., devotee of quantitative methods, statistics geek). I took a workshop and learned to program a Burroughs 205 computer (competitor with the IBM 650); the 205 took up an entire floor of Nebraska Hall, which had to have special air conditioning installed to accommodate the heat generated by the behemoth. My job was to take randomly generated judgmental data matrices and convert them into a matrix of cosines of angles of separation among vectors representing stimulus objects. It took me six months to create and test the program; on today's equipment, it would require a few hours. Bob took over the resulting matrix and extracted latent roots to be compiled into empirical sampling distributions.

The work was in the tradition of metric scaling invented by Thurstone and generalized to the multidimensional case by Richardson and Torgerson and others; it was heady stuff. I was allowed to operate the computer in the middle of the night, bringing it up and shutting it down by myself. Bob found an office for me to share with a couple of graduate students in Ed Psych. I couldn't believe my good luck; from scrubbing floors to programming computers almost overnight. I can recall virtually every detail of those two years I spent working for Bob, first on the MDS project, then on a few other research projects he was conducting (even creating Skinnerian-type programmed instruction for a study of learner activity; my assignment was to program instruction in the Dewey Decimal system).

Stake was an attractive and fascinating figure to a young man who had never in his 20 years on earth traveled farther than 100 miles from his birthplace. He drove a Chevy station wagon, dusty rose and silver. He lived on the south side of Lincoln, a universe away from the lower-middle class neighborhoods of my side of town. He had a beautiful wife and two quiet, intense young boys who hung around his office on Saturdays silently playing games with paper and pencil. In the summer of 1961, I was invited to the Stake's house for a barbecue. Several graduate students were there (Chris Buethe, Jim Beaird, Doug Sjogren). The backyard grass was long and needed mowing; in the middle of the yard was a huge letter "S" carved by a lawn mower. I imagined Bernadine having said once too often, "Bob, would you please mow the backyard?" (Bob's children tell me that he was accustomed to mowing mazes in the yard and inventing games for them that involved playing tag without leaving the paths.)

That summer, Bob invited me to drive with him to New York City to attend the ETS Invitational Testing Conference. Bob's mother would go with us. Mrs. Stake was a pillar of the small community, Adams, 25 miles south of Lincoln where Bob was born and raised. She regularly spoke at auxiliary meetings and other occasions about the United Nations, then only 15 years old. The trip to New York would give her a chance to renew her experiences and pick up more literature for her talks. Taking me along as a spare driver on a 3,500 mile car trip may not have been a completely selfless act on Bob's part, but going out of the way to visit the University of Wisconsin so that I could meet Julian Stanley and learn about graduate school definitely was generous. Bob had been corresponding with Julian since the Spring of 1961. The latter had written his colleagues around the country urging them to test promising young students of their acquaintance and send him any information about high scores. In those pre-GRE days, the Miller Analogies Test and the Doppelt Mathematical Reasoning Test were the instruments of choice. Julian was eager to discover young, high scorers and accelerate them through a doctoral program, thus preventing for them his own misfortune of having wasted four of his best years in an ammunition dump in North Africa during WW II--and presaging his later efforts to identify math prodigies in middle school and accelerate them through college. Bob had created his own mental ability test, named with the clever pun QED, the Quantitative Evaluative Device. Bob asked me to take all three tests; I loved taking them. He sent the scores to Julian, and subsequently the stop in Madison was arranged. Bob had made it clear that I should not attend graduate school in Lincoln.

We drove out of Lincoln--the professor, the bumpkin and the UN Ambassador--on October 27, 1961. Our first stop was Platteville, Wisconsin, where we spent the night with Bill Jensen, a former student of Bob's from Nebraska. Throughout the trip we were never far from Bob's former students who seemed to feel privileged to host his retinue. On day two, we met Julian in Madison and had lunch at the Union beside Lake Mendota with him and Les McLean and Dave Wiley. The company was intimidating; I was certain that I did not fit in and that Lincoln was the only graduate school I was fit for. We spent the third night sleeping in the attic apartment of Jim Beaird, whose dissertation that spring was a piece of the Stake MDS project; he had just started his first academic job at the University of Toledo. The fourth day took us through the Allegheny Mountains in late October; the oak forests were yellow, orange and crimson, so unlike my native savanna. We shared the driving. Bob drove through rural New Jersey searching for the small community where his brother Don lived; he had arranged to drop off his mother there. The maze was negotiated without the aid of road maps or other prostheses; indeed, none was consulted during the entire ten days. That night was spent in Princeton. Fred Kling, a former ETS Princeton Psychometric Fellow at Princeton with Bob, and his wife entertained us with a spaghetti dinner by candlelight. It was the first time in my life I had seen candles on a dinner table other than during a power outage, as it was also the first time I had tasted spaghetti not out of a can.

The next day we called on Harold Gulliksen at his home. Gulliksen had been Bob's adviser at Princeton. We were greeted by his wife, who showed us to a small room outside his home office. We waited a few minutes while he disengaged from some strenuous mental occupation. Gulliksen swept into the room wearing white shirt and tie; he shook my hand when introduced; he focused on Bob's MDS research. The audience was over within fifteen minutes. I didn't want to return to Princeton.

We drove out to the ETS campus. Bob may have been gone for three years, but he was obviously not forgotten. Secretaries in particular seemed happy to see him. Bob was looking for Sam Messick. I was overwhelmed to see that these citations--(Abelson and Messick, 1958)--were actual persons, not like anything I had ever seen in Nebraska of course, but actual living, breathing human beings in whose presence one could remain for several minutes without something disastrous happening. Bob reported briefly on our MDS project to Messick. Sam had a manuscript in front of him on his desk. "Well, it may be beside the point," Messick replied to Bob's description of our findings. He held up the manuscript. It was a pre-publication draft of Roger Shepard's "Analysis of Proximities," which was to revolutionize multidimensional scaling and render our monte carlo study obsolete. It was October 30, 1961. It was Bob Stake's last day on earth as a quantoid.

The ETS Invitational Testing Conference was held in the Roosevelt Hotel in Manhattan. We bunked with Hans Steffan in East Orange and took the tube to Manhattan. Hans had been another Stake student; he was a native German and I took the opportunity to practice my textbook Deutsch. I will spare the reader a 21-year-old Nebraska boy's impressions of Manhattan, all too shopworn to bear repeating. The Conference was filled with more walking citations: Bob Ebel, Ledyard Tucker, E. F. Lindquist, Ted Cureton, famous name after famous name. (Ten years later, I had the honor of chairing the ETS Conference, which gave me the opportunity to pick the roster of speakers along with ETS staff. I asked Bob to present his ideas on assessment; he gave a talk about National Assessment that featured a short film that he had made. People remarked that they were not certain that he was being "serious." His predictions about NAEP were remarkably prescient.)

We picked up Bob's mother in Harrisburg, Pennsylvania, for some reason now forgotten. While we had listened to papers, she had invaded and taken over the U.N. We pointed the station wagon west; we made one stop in

Toledo to sleep for a few hours. I did more than my share behind the wheel. I was extremely tired, having not slept well in New York. Bob and I usually slept in the same double bed on this trip and I was too worried about committing some gross act in my sleep to rest comfortably. I had a hard time staying awake during my stints at the wheel, but I would not betray weakness by asking for relief. I nearly fell asleep several times through Ohio, risking snuffing out two promising academic careers and breaking Adams, Nebraska's only diplomatic tie to the United Nations.

To help relieve the boredom of the long return trip, Bob and I played a word game that he had learned or invented. It was called "Ghost." Player one thinks of a five-letter word, say "spice." Player two guesses a five-letter word to start; suppose I guessed "steam." Player one superimposes, in his mind, the target word "spice" and my first guess "steam" and sees that one letter coincides--the "s." Since one letter is an odd number of letters, he replies "odd." If no letters coincide he says "even." If I had been very lucky--actually unlucky--and first guessed "slice," player one would reply "even" because four letters coincide. (This would actually have been an unlucky start since one reasonably assumes that the initial response "even" means that zero letters coincide. I think that games of this heinous intricacy are not unknown to Stake children.) Through a process of guessing words and deducing coincidences from "odd" and "even" responses, player two eventually discovers player one's word. It is a difficult game and it can consume hundreds of miles on the road. Several rounds of the game took us through Ohio, Indiana, Illinois. Somewhere around the Quad Cities, Bob played his trump card. He was thinking of a word that resisted all my most assiduous attempts at deciphering. Finally, outside Omaha I conceded defeat. His word was "ouija," as in the board. Do we take this incident as in some way a measure of this man?

By the time I arrived in Lincoln, a Western Union Telegram from Julian was waiting. I had never before received a telegram--or known anyone who had. I was flattered; I was hooked. Three months later, January 1962, I left Lincoln, Stake and everything I had known my entire life for graduate school. Bob and I corresponded regularly during the ensuing years. He wrote to tell me that he had taken a job at Urbana. I told him I was learning all that was known about statistics. He wrote several times during his summer, 1964, at Stanford in the institute that Lee Cronbach and Richard Atkinson conducted. Clearly it was a transforming experience for him. I was jealous. When I finished my degree in 1965, Bob had engineered a position for me in CIRCE at Univ. of Illinois. I was there when Bob wrote his "Countenance" paper; I pretended to understand it. I learned that there was a world beyond statistics; Bob had undergone enormous changes intellectually since our MDS days. I admired them, even as I recognized my own inability to follow. I spent two years at CIRCE; I think I felt the need to shine my own light away from the long shadows. I picked a place where I thought I might shine: Colorado.

Bob and I saw very little of each other from 1967 on. In the early 1970s, I invited him to teach summer school at Boulder. He gave a seminar on evaluation and converted all my graduate students into Stake-ians. But I saw little of him that summer. We didn't connect again until 1978.

When the year 1978 arrived, I was at the absolute height of my powers as a quantoid. My book on time-series experiment analysis was being reviewed by generous souls who called it a "watershed." Meta-analysis was raging through the social and behavioral sciences. I had nearly completed the class-size meta-analysis. The Hastings Symposium, on the occasion of Tom Hastings's retirement as head of CIRCE, was happening in Urbana in January. I attended. Lee Cronbach delivered a brilliant paper that gradually metamorphosed into his classic *Designing Evaluations of Educational and Social Programs*. Lee argued that the place of controlled experiments in educational evaluation is much less than we had once imagined. "External validity," if we must call it that, is far more important than "internal validity," which is after all not just an impossibility but a triviality. Experimental validity can not be reduced to a catechism. Well, this cut to the heart of my quantoid ideology, and I remember rising during the discussion of Lee's paper to remind him that controlled, randomized experiments worked perfectly well in clinical drug trials. He thanked me for divulging this remarkable piece of intelligence.

That summer I visited Eva Baker's Center for the Study of Evaluation at UCLA for eight weeks. Bob came for two weeks at Eva's invitation. One day he dropped a sheet of paper on my desk that contained only these words:

Chicago	6
New York	5
Lincoln	6
Phoenix	8
Urbana	10
San Francisco	10

We were back to ghost, I could tell. I worked all day and half the night on it. I was stuck. Then I remembered that he was staying by himself in a bare apartment just off campus. When I visited it several days before, there had only been a couch, a phone and a phonebook in the living room. I grabbed a phonebook and started perusing it. There near the front was a list of city names and area codes: Chicago 312, New York 212, Lincoln 402; 3+1+2=6, 2+1+2=5, 4+0+2=6, etc. Bingo! He didn't get me this time.

I was a quantoid, and "what I do best" was peaking. I gave a colloquium at Eva's center on the class size meta-analysis in mid-June. People were amazed. Jim Popham asked for the paper to inaugurate his new journal *Educational Evaluation and Policy Analysis*. He was welcome to it.

June 30, 1978, dawned inauspiciously; I had no warning that it would be my last day on earth as a quantoid. Bob was to speak at a colloquium at the Center on whatever it was that was on his mind at that moment. Ernie House was visiting from Urbana. I was looking forward to the talk, because Bob never gave a dull lecture in his life. That day he talked about portrayal, complexity, understanding; qualities that are not yet nor may never be quantities; the ineffable (Bob has never been a big fan of the "effable"). I listened with respect and admiration, but I listened as one might listen to stories about strange foreign lands, about something that was interesting but that bore no relationship to one's own life. Near the end when questions were being asked I sought to clarify the boundaries that contained Bob's curious thoughts. I asked, "Just to clarify, Bob, between an experimentalist evaluator and a school person with intimate knowledge of the program in question, who would you trust to produce the most reliable knowledge of the program's efficacy?" I sat back confident that I had shown Bob his proper place in evaluation--that he couldn't really claim to assess impact, efficacy, cause-

and-effect with his case-study, qualitative methods--and waited for his response, which came with uncharacteristic alacrity. "The school person," he said. I was stunned. Here was a person I respected without qualification whose intelligence I had long admired who was seeing the world far differently from how I saw it.

Bob and Ernie and I stayed long after the colloquium arguing about Bob's answer, rather Ernie and I argued vociferously while Bob occasionally interjected a word or sentence of clarification. I insisted that causes could only be known (discovered, found, verified) by randomized, controlled experiments with double-blinding and followed up with statistical significance tests. Ernie and Bob argued that even if you could bring off such an improbable event as the experiment I described, you still wouldn't know what caused a desirable outcome in a particular venue. I couldn't believe what they were saying; I heard it, but I thought they were playing Jesuitical games with words. Was this Bob's ghost game again?

Eventually, after at least an hour's heated discussion I started to see Bob and Ernie's point. Knowledge of a "cause" in education is not something that automatically results from one of my ideal experiments. Even if my experiment could produce the "cause" of a wonderful educational program, it would remain for those who would share knowledge of that cause with others to describe it to them, or act it out while they watched, or somehow communicate the actions, conditions and circumstances that constitute the "cause" that produces the desired effect. They--Bob and Ernie--saw the experimenter as not trained, not capable of the most important step in the chain: conveying to others a sense of what works and how to bring it about. "Knowing" what caused the success is easier, they believed, than "portraying" to others a sense for what is known.

I can not tell you, dear reader, why I was at that moment prepared to accept their belief and their arguments, but I was. What they said in that hour after Bob's colloquium suddenly struck me as true. And in the weeks and months after that exchange in Moore Hall at UCLA, I came to believe what they believed about studying education and evaluating schools: many people can know causes; few experiments can clarify causal claims; telling others what we know is the harder part. It was my last day on earth as a quantoid.

In the early 1970s, Bob introduced me to the writings of another son of Lincoln, Loren Eiseley, the anthropologist, academic and author, whom Wystan H. Auden once named as one of the leading poets of his generation. Eiseley wrote often about his experiences in the classroom; he wrote of "hidden teachers," who touch our lives and never leave us, who speak softly at the back of our minds, who say "Do this; don't do that."

In his book *The Invisible Pyramid,* Eiseley wrote of "The Last Magician." "Every man in his youth--and who is to say when youth is ended?--meets for the last time a magician, a man who made him what he is finally to be" (p. 137). For Eiseley, that last magician is no secret to those who have read his autobiography, *All the Strange Hours*; he was Frank Speck, an anthropology professor at the University of Pennsylvania who was Eiseley's adviser, then colleague, and to whose endowed chair Eiseley succeeded upon Speck's retirement. (It is a curious coincidence that all Freudians will love that Eiseley's first published book was a biography of Francis Bacon entitled *The Man Who Saw Through Time*; Francis Bacon and Frank Speck are English and German translations of each other.)

Eiseley described his encounter with the ghost of his last magician:

> "I was fifty years old when my youth ended, and it was, of all unlikely places, within that great unwieldy, structure built to last forever and then hastily to be torn down--the Pennsylvania Station in New York. I had come in through a side doorway and was slowly descending a great staircase in a slanting shaft of afternoon sunlight. Distantly I became aware of a man loitering at the bottom of the steps, as though awaiting me there. As I descended he swung about and began climbing toward me.
>
> "At the instant I saw his upturned face my feet faltered and I almost fell. I was walking to meet a man ten years dead and buried, a man who had been my teacher and confidant. He had not only spread before me as a student the wild background of the forgotten past but had brought

> alive for me the spruce-forest primitives of today. With him I had absorbed their superstitions, handled their sacred objects, accepted their prophetic dreams. He had been a man of unusual mental powers and formidable personality. In all my experience no dead man but he could have so wrenched time as to walk through its cleft of darkness unharmed into the light of day.
>
> "The massive brows and forehead looked up at me as if to demand an accounting of that elapsed time during which I had held his post and discharged his duties. Unwilling step by step I descended rigidly before the baleful eyes. We met, and as my dry mouth strove to utter his name, I was aware that he was passing me as a stranger, that his gaze was directed beyond me, and that he was hastening elsewhere. The blind eye turned sidewise was not, in truth, fixed upon me; I beheld the image but not the reality of a long dead man. Phantom or genetic twin, he passed on, and the crowds of New York closed inscrutably about him" (pp. 137-8).

Eiseley had seen a ghost. His mind fixed on the terror he felt at encountering Speck's ghost. They had been friends. Why had he felt afraid?

> "On the slow train running homeward the answer came. I had been away for ten years from the forest. I had had no messages from its depths. . . . I had been immersed in the postwar administrative life of a growing university. But all the time some accusing spirit, the familiar of the last wood-struck magician, had lingered in my brain. Finally exteriorized, he had stridden up the stair to confront me in the autumn light. Whether he had been imposed in some fashion upon a convenient facsimile or was a genuine illusion was of little importance compared to the message he had brought. I had starved and betrayed myself. It was this that had brought the terror. For the first time in years I left my office in midafternoon and sought the sleeping silence of a nearby cemetery. I was as pale and drained as the Indian pipe plants without chlorophyll that rise after rains on the forest floor. It was time for a change. I wrote a letter and studied timetables. I was returning to the land that bore me" (p. 139).

Whenever I am at my worst--rash, hostile, refusing to listen, unwilling even to try to understand--something tugs at me from somewhere at the back of consciousness, asking me to be better than that, to be more like this person or that person I admire. Bob Stake and I are opposites on most dimensions that I can imagine. I form judgments prematurely; he is slow to judge. I am impetuous; he is reflective. I talk too much; perhaps he talks not enough. I change my persona every decade; his seemingly never changes. And yet, Bob has always been for me a hidden teacher.

References

Eiseley, Loren (1970). *The invisible pyramid*. New York: Scribner.

Eiseley, Loren (1975). *All the strange hours*. New York: Scribner.

Bob Stake Meets Mr. Rogers

David E. Balk
Oklahoma State University[1]

> "I'm glad you've chosen to go to the University of Illinois, David," the Head of the Department of Educational Psychology at Arizona State University said to me in May of 1978.
>
> "Why is that, Dr. Von Wagenen?" the clueless graduate-student-in-the-making asked in reply.
>
> "Because now you'll get to meet Bob Stake," the Department Head said.
>
> "Who is Bob Stake and why should I want to meet him?" was my tactless question.
>
> "Bob Stake has made the most important contributions to epistemology of anyone in education over the past twenty years," was his answer.

Now it is twenty years almost to the day since Keith Von Wagenen provided this succinct, highly accurate, richly evocative description of Bob Stake--and--what I did not know--forecast the influence that meeting Bob Stake would have upon me.

My years at the University of Illinois were deeply formative. The culture of this marvelous institution, the quiet but persistent expectations to achieve a standard exceeding what others had thought excellent, and the opportunities to grow by listening and reading and contributing all had an impact upon me. I had the chance to work closely with a few persons--Helen Farmer most especially--and the great opportunity to know two individuals whose presence in my life has had a lasting impact to this day: Tom Hastings and Bob Stake.

I came to the University of Illinois to get a Ph.D. in Counseling Psychology, which I did in 1981, but I ended up majoring in CIRCE with side journeys to the philosophy of science. The very chance to organize my time at the U of I in this fashion I owe in some part to the patience and forbearance of my major professor, M. Jean Phillips. On more than one occasion Jean Phillips noted that she had never seen me counsel anyone during my two and one-half years working for my doctoral degree in Counseling Psychology. I took courses in the philosophy of science and became fascinated by the minds and personalities in CIRCE. I became attracted to the conversations that were possible at any moment by just crossing the hall from my cubicle on the second floor of the College of Education to visit with Tom Hastings or Bob Stake or Ernie House or Oli Proppe or Deborah Trumbull or to attend the brown bag lunches which as I remember were held on Thursdays at Noon.

---

[1] Department of Family Relations and Child Development, College of Human Environmental Sciences Stillwater, OK 74078

Please don't misunderstand me. My work with the Counseling Psychology faculty at the University of Illinois was of importance to me and has had a lasting impact in my life. They introduced me to life-span developmental psychology and showed me how to integrate that point of view into strategies to help persons at risk. I have become known in some scholarly circles for bereavement research with adolescents, and those efforts began with my dissertation written in the College of Education and with the wonderful opportunities I had to meet with Helen Farmer and Jean Phillips as well as with Lenore Harmon and Jim Wardrop.

Yet when I think of how I changed while at the University of Illinois, I turn immediately to CIRCE and to the two persons I most associate with CIRCE: Tom Hastings and Bob Stake. Over the past many years I have had numerous opportunities to make use of the thinking and the writings of these two men. They influenced some of my work in program evaluation at a community mental health center in Tucson, AZ. Their thinking and writing took center stage when I began teaching a graduate course in program evaluation at Kansas State University. My work in that course at K-State led me to the idea for this paper and its title: "Bob Stake Meets Mr. Rogers."

That course in program evaluation evolved as I worked from year to year to figure out how to get graduate students to grasp the issues and ideas central to program evaluation. I didn't begin the course with Bob Stake and responsive evaluation, but rather with the question "What does it mean to evaluate a program?" Students were told they would have to complete an evaluation of some program by the end of the semester, and most typically there were several programs on the campus or in the community that welcomed the chance to have an evaluation of their efforts. There was no textbook but rather a lengthy list of books and articles on reserve in the K-State library from which students were to select and write thought papers of 5-8 pages in length throughout the course of the semester. It did not escape the students' attention that there was a considerable amount of ambiguity to this course. Borrowing a phrase from the philosophy of science, I told them learning about program evaluation was akin to having to build your ship while already being at sea.

When I lectured, I noticed that I did so just sitting there with the students and giving them overviews of Tyler, of Hastings, of Stufflebeam, of Scriven, and of Stake. I never gave them an overview of Rossi (although his books were placed on reserve and I noted for the students that the Rossi & Freeman textbook apparently was the most widely used textbook in evaluation courses). I found myself more and more convinced that the most persuasive approach to conducting program evaluations was qualitative, and I told students to take my comments with several grains of salt since I was biased toward what I considered the CIRCE connection in my life.

One day while reading in the Shadish et. al. book Foundations of Program Evaluation, I came across this statement about Bob Stake: ". . . his early teaching was in training school counselors" (Shadish et. al., 1991, 272). This information was an insight into what had up to that time remained unspoken and probably unformed in my understanding of Bob Stake's work. Now I did not go running naked from my office yelling "Eureka" to faculty and students on the K-State campus. I just started pondering some more and thought I found a possible way to explain an influence on Bob Stake's thinking that before I would not have considered. That influence came from my own discipline--you remember, counseling psychology--and could be attributed only to one source: Carl Rogers. Thus, my paper's title "Bob Stake Meets Mr. Rogers."

From the early 1940s into the 1960s Rogers revolutionized thinking in counseling psychology with such works as *Counseling and Psychotherapy* (Rogers, 1942), Client-

Centered Therapy (Rogers, 1951), *Psychotherapy and Behavior Change* (Rogers & Dymond, 1954), and *On Becoming a Person* (Rogers, 1961). Remarkable changes occurred in counseling circles when Carl Rogers began to assert his ideas and to attack the behaviorist positions that had seemed intractable until Rogers published his writings (Garth J. Blackham, personal communication, sometime in 1975).

Thus, when I read the statement that Bob Stake had taught school counselors, I figured it had to have been in the 1950s, and I knew the major influence upon counseling at that time was Carl Rogers. I have thought about asking Bob his own views on this matter, and this paper is one way of asking him to respond.

Let me set forth my ideas about this matter. What does Bob say about the efforts of a program evaluator? He says, "When you hire an evaluator, you aren't hiring a person who has a great deal of wisdom about your problems. You aren't going to get someone who will capture a truth that is really crucial to your program. It is much more likely that whatever truths, whatever solutions there are, exist in the minds of the people who are running the program, those participating in the program, those patrons of the program. . . . (The evaluator) is making his greatest contribution, I think, when he is helping people discover ideas, answers, solutions, within their own minds" (Stake, 1975a, 36).

Carl Rogers insisted that clients know what is in their best interests and insisted that the only change possible for clients has to come from within themselves. In Rogers' thinking, clients make changes in their lives when the counselor

- Relates to them authentically,
- Demonstrates empathetic understanding of their situations,
- Learns their frame of reference, and
- Manifests unconditional positive regard for them.

The conditions for change emerge because of the responsive nature of the counselor-client relationship. In Carl Rogers' words, personal growth occurs "when the client perceives, to a minimal degree, the genuineness of the counselor and the acceptance and empathy which the counselor experiences for him" (Rogers, 1967a, 96).

Is there any means to tie these ideas to Stake? Well, I think the answer is yes. For one thing, Stake says the dominant purpose for evaluation is to be of service, and such service is obstructed unless evaluators learn the interests and language of their audiences. Responsive evaluators couch their reports in the language of the persons in and around the program. This tack seems very much like client-centered counselors learning the frames of reference of their clients.

Bob Stake has said qualitative research underscores the value of experience, and has emphasized that qualitative researchers attempt to evoke empathetic understanding of others' experience. Responsiveness as an evaluator requires openness to and acceptance of the experience of others. Rogers made openness to and acceptance of the experiences of others a necessary attribute of effective counselors.

The connection between Stake and Rogers goes deeper than their commitment to being responsive to clients or programs. The connection involves their very understanding of scientific inquiries into human endeavors.

Carl Rogers expressed both concern and amazement at the positivist approach that had overtaken the behavioral sciences. He wrote, "I object to the process of depersonalization and dehumanization of the individual in our culture. I regret that the behavioral sciences seem to me to be promoting and reinforcing this trend" (Rogers, 1968, 59). He protested vigorously that the rigid determinism advocated by B. F. Skinner truncated human experience by leaving out volition and intentionality (see, for instance, Rogers, 1967b).

What does Bob Stake champion in the effort to learn more about human endeavors? Well, first of all he makes a profound argument for the value of case study research and for the merit of knowing the single case in rich detail. He notes, for instance, in *The Art of Case Study Research* that "We study a case when it itself is of very special interest. We look for the detail of interaction with its contexts. Case study is the study of the particularity and complexity of a single case, coming to understand its activity within important circumstances" (Stake, 1995, xi). Stake's valuing of the single case is quite like Carl Rogers' valuing of the individual.

Second, Bob Stake argues for the value of achieving understanding in contrast to providing explanation. He has developed sophisticated, persuasive arguments (referencing such philosophers as Michael Polanyi and Georg Henrik von Wright) that the complexity of the human world--let's say, for instance, the complexity of an educational program resists our efforts to control all the realities involved. Stake makes explicit reference to human intentionality and quotes von Wright's statement that understanding is connected with intentionality and with planning and aims and purposes (Stake, 1995).

Carl Rogers asks a question posed also by Bob Stake. The question is "How do we know?" Rogers believes in formalistic research designs using hypotheses, adequate testing, sophisticated research designs, precision, and statistical methodology; he said so in his paper "Some Thoughts Regarding the Current Presuppositions of the Behavioral Sciences" (Rogers, 1968). Rogers systematically examined the effects of his psychotherapy with schizophrenic patients (Rogers, 1967c).

However, Rogers primarily believes in naturalistic generalizations made from highly personal information. Thus, rather than an appeal to science as a means to answer the question "How do we know?," Rogers said "The more one pursues this question, the more one is forced to realize that in the last analysis, knowledge rests on the subjective: I experience. . . . All knowledge, including all scientific knowledge, is a vast inverted pyramid resting on this tiny, personal, subjective base. . . . I think that it is not too much to say that knowing, even in the hardest sciences, is a risky, uncertain, subjective leap even when it is most 'objective.' We do no one a service by pretending it is not this" (Rogers, 1968, 60).

What has Bob Stake said about knowledge and how we know? Well, first of all he has made a clear distinction between two approaches to gaining knowledge:

- the hypothetico-deductive method using carefully controlled research designs leading to formalistic generalizations and

- an inductive approach giving credence to the unexpected and the uncontrolled and leading to naturalistic generalizations.

He has said both approaches have value: the former offers precision and objectivity, and the latter offers insight and understanding. He has said the much more common way of gaining knowledge is through observation leading to naturalistic generalizations (Stake, 1986; Stake & Trumbull, 1982). He has said program evaluators should use rigorous methods that enable stakeholders to gain knowledge of their programs by forming conclusions based on rich representations of what the evaluator has observed. For example, consider the following words from Bob Stake:

> To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations. He arranges for various persons to observe the program, and with their help prepares brief narratives, portrayals, product displays, graphs, etc. He finds out what is of value to his audiences, and gathers expression of worth from various individuals whose points of view differ. Of course, he checks the quality of his records; he gets program personnel to react to the accuracy of his portrayals; authority figures to react to the importance of various findings; and audience members to react to the relevance of his findings (Stake, 1975b, 14).

Finally, both Carl Rogers and Bob Stake admitted they overstated their own cases in order to get a firm foothold in disciplines. Berenson and Carkhuff (1967) suggested such was the case with Rogers. Stake admitted such was the case with himself:

> I see it as unfortunately necessary to overstate the distinction between academic research and practical inquiry as a step toward improving and legitimizing inquiries that are needed for understanding and problem solving but which are unlikely to produce vouchsafed generalizations (Stake, 1978, 7).
>
> The responsive evaluator is guided largely by the particular situation. How much to emphasize the particular or the general is a relative matter. Of course,

> there will be the day when I will say, 'We went too far. . . (Stake, 1975a, 34).

It can't be said that either Carl Rogers or Bob Stake has won the day in his own field. As for Rogers, various research studies have indicated that positive outcomes for clients are associated with the conditions Rogers termed necessary for personal growth. However, cognitive or symbolic mediational processes as much as or more than affective elements lead to client change. Several studies have identified important client gains attributable not merely to empathetic understanding and unconditional positive regard but to the direct instructions and influence of the counselor (Blackham, 1975).

And as for Bob Stake, let me offer the following anecdotes.

> I have been a faculty member in two colleges whose types are found only I am sure in land grant institutions. At K-State my college was called the College of Human Ecology, and at Oklahoma State University it is called the College of Human Environmental Sciences. At Illinois it is the College of Agricultural, Consumer, and Environmental Sciences. At neither K-State's College of Human Ecology nor at Oklahoma State University's College of Human Environmental Sciences is Bob Stake's name or his work well known. Whereas qualitative methodology is becoming more accepted in these colleges, people would be prone to ask a question stated early in this talk, "Who is Bob Stake and why should I want to meet him?"
>
> A version of this very question was uttered last December when I was giving some lectures at Colorado State University. I got into a conversation with a faculty member who taught program evaluation and who knew I had taught program evaluation; he wanted to know what textbook I had used. I told him I had not used a text but rather had made available a legion of material in the library, and I said I had become more and more attracted to Bob Stake's approach. He said something to the effect of "Who is Bob Stake?" And he informed me he used the Rossi & Freeman text.
>
> About eight years ago I was a member of the American Psychological Association division that is devoted to

> psychological measurement. Division members were invited to nominate distinguished speakers to address the annual conference, and I nominated Bob after calling him to see if he would accept such an invitation if it were given. I had to give the Program Chair for the division a brief overview of Bob's accomplishments, and I stressed his work in forging a new appreciation for qualitative data and vicarious generalizations and case study research. The response of the program chairperson to my nomination of Bob is hard to forget. The division decided not to invite Bob "because his thinking is not in the main with the rest of the members of this division." I believe my response was something to the effect of "Well, isn't that all the more reason to invite him?," but I may be guilty of delusions of grandeur as I recreate this event.

To end my talk, I want to do a few things. First, let me repeat the title of this talk, which is "Bob Stake Meets Mr. Rogers." And then I want to furnish you with two quotes. Who said these words: Bob Stake or Carl Rogers?

> "I believe I am accurate in saying that educators too are interested in learnings which make a difference. Simple knowledge of facts has its value. To know who won the battle of Poltava, or when the umpteenth opus of Mozart was first performed, may win $64,000 or some other sum for the possessor of this information, but I believe educators in general are a little embarrassed by the assumption that the acquisition of such knowledge constitutes education. Speaking of this reminds me of a forceful statement made by a professor of agronomy in my freshman year in college. Whatever knowledge I gained in his course has departed completely, but I remember how, with World War I as his background, he was comparing factual knowledge with ammunition. He wound up his little discourse with the exhortation, 'Don't be a damned ammunition wagon; be a rifle!' I believe most educators would share this sentiment that knowledge exists primarily for use."[2]

And who said these words: Bob Stake or Carl Rogers?

---

[2] Rogers, 1961, page 281.

> "The fisherman examines not only the size of the catch but also the holes in the net."[3]

Well, this is the end of my talk. Like Keith von Wagenen's wish for me, I am glad that I chose to come to study at the University of Illinois. You see, by doing that I got to meet Bob Stake. And now I know why I always wanted to meet him, even back when I didn't know who he was.

---

[3] Old Nebraska proverb written on the CIRCE chalkboard and seen by the author sometime during his Ph.D. education at the University of Illinois.

References

Berenson, B. G. & Carkhuff, R. R. (Eds.). (1967). *Sources of gain in counseling and psychotherapy.* New York: Holt.

Blackham, G. J. (1975). *Counseling: Theory, process, and practice.* Belmont, CA: Wadsworth.

Rogers, C. R. (1942). *Counseling and psychotherapy: Newer concepts in practice.* Boston: Houghton Mifflin.

Rogers, C. R. (1951). *Client-centered therapy: Its current practice, implications, and theory.* Boston: Houghton Mifflin.

Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy.* Boston: Houghton Mifflin.

Rogers, C. R. (1967a). The interpersonal relationship: The core of guidance. In C. R. Rogers & B. Stevens (Eds.), *Person to person: The problem of being human. A new trend in psychology* (pp. 89-103). Lafayette, CA: Real People Press.

Rogers, C. R. (1967b). Learning to be free. In C. R. Rogers & B. Stevens (Eds.), *Person to person: The problem of being human. A new trend in psychology* (pp. 47-66). Lafayette, CA: Real People Press.

Rogers, C. R. (1967c). Some learnings from a study of psychotherapy with schizophrenics. In C. R. Rogers & B. Stevens (Eds.), *Person to person: The problem of being human. A new trend in psychology* (pp. 181-192). Lafayette, CA: Real People Press.

Rogers, C. R. (1968). Some thoughts regarding the current propositions of the behavioral sciences. In W. R. Coulson & C. R. Rogers (Eds.), *Man and the science of man* (pp. 54-72). Columbus, OH: Charles E. Merrill.

Rogers, C. R. & Dymond, R. F. (Eds). (1954). *Psychotherapy and personality change: Co-ordinated research studies in the client-centered approach.* Chicago: University of Chicago Press.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.

Stake, R. E. (1975a). An interview with Robert Stake on responsive evaluation. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 33-38). Columbus, OH: Charles E. Merrill.

Stake, R. E. (1975b). To evaluate an arts program. In R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 13-31). Columbus, OH: Charles E. Merrill.

Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher, 7,* 5-8.

Stake, R. E. (1986). An evolutionary view of educational improvement. In E. House (Ed.), *New directions in educational evaluation* (pp. 89-102). London: Farmer.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R. E. & Trumbull, D. J. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Science, 7,* 1-12.

Tom and Bob:
CIRCE '64 to '67:
Evaluation Sweetwater on the Illinois Plains:
Portrait of an Education:
A Responsive Reflection:
Five Colons in Search of a Paper

Thomas O. Maguire
The University of Alberta

To seek the sweet water, we must look beyond the Boneyard (which was not sweet by any stretch of the imagination), and look back to 270 Education. This is CIRCE, home of responsive evaluation, the well spring of revolutionary evaluation thought. We open the door. Lois Williamson. She of the Effingham accent and sensible shoes signifies that here, evaluation will be well grounded and based on common sense. No West Coast weirdness, no eastern PERT, only solid, corn-fed insight. Lois Williamson, can she be one who calls sailors onto the rocks of quasi-experimental approaches? I think not. She is the guardian of the Gods (as well as keeper of the keys to the kingdom (i.e. the keys to the thermo fax machine)). Principal among the gods is J. Thomas Hastings, Director of the Illinois Statewide Testing Program, and the soul of CIRCE. He is "Tom" to all who know him (except Miss Williamson who always refers to him as Mr. Hastings). Close your eyes and recall a lean man with a brushcut and smile. He leans back in his secretarial chair, lights a cigarette, and consumes it with powerful draws. Running his hand over his head, he might say about the education of evaluators [and I've cribbed this from a set of interviews that were done by Gabe and Connie Della Piana],

> "Well, you know, I want to try and leave a good bit of leeway for the student to wander off on some of his own concepts and some of his own ways of implementing certain basic premises. I think that is the only way that we are going to grow. Now some people have accused me because of that, oh they call it lots of things, from tom foolery to teaching nothing really, just mentioning a few things that people have done or something. Well it was never my

> intent, and I don't think it was my practice, but rather to get them . . . it is one reason in my, and I've mentioned it several times now, but the so-called theory of advanced evaluation seminar, it is one reason that I always tried to have a hands on project going, and I didn't expect them to all come out the same or with the same approaches to data gathering or same approaches to interpretation, but I did expect them to come out with data gathering and with interpretations they could support, they could show evidence for. I was more interested in their inventiveness in finding alternative ways than I was in their doing it only one way. Now I know over at Ohio State when Dan Stufflebeam was still there, before he went to Western Michigan University, I heard directly from quite a few of his students that Dan taught, and this would be advanced courses, but CIPP, the one that he and several others invented, that was the way you do when you evaluate programs period. Oh he might on the way through mention something else such as a short paper by Bob Stake on the responsive evaluation approach, but not as something they should follow, but they should be aware that this is a way that someone had looked at it. Well I'd find it abhorrent to pick out some way to do it and say here is the way. So that's my definition of it. I don't really care a lot how many happen to agree, and I don't remember others using those words, but I certainly heard phrases from them which would indicate that they too thought there was this difference between training and evaluation."

This philosophical orientation is part of the mountain snowpack that produces the evaluation sweet water. At CIRCE, education was a process of exploration.

Down the hall, we find Robert E. Stake. Bob to all who know him (except, again Miss Williamson who refers to him (when necessary) as Mr. Stake.) As we will soon learn, Bob and Tom are a study in contrasts. Bob doesn't smoke, doesn't have a brushcut, and doesn't make small talk. But he does take grape nuts on his ice cream. If Tom is the soul of CIRCE, then Bob is the creative intellect.

Further down the hall is a third office that will eventually be occupied by the Wisconsin cherub, Gene Glass... a teenager from Nebraska who studied at the Laboratory of Experimental Design under the great evangelist Julian Stanley. He, Gene, is said to know a great deal about alpha factor analysis of fallible variables, something that I am confident will enhance the quality of my future life. Gene is master at debate. But in arguments about matters that are monumentally unimportant, he will never be able to outmaneuver Peter Taylor, the first CIRCE doctoral student.

Speaking of CIRCE metaphorically (and we retrospective sweet water practitioners like to do that a lot), If Bob was the ego, Tom was the super ego, was Gene the id?-- Perhaps not, Gene was a superb mentor. The way he guided aspects of my own intellectual growth was very much appreciated both at that time and in retrospect.

I learned a lot from my mentors. Long before "think aloud" procedures were refined as a way of helping to understand the inner workings of problem solvers; Tom Hastings was actively demonstrating the skill on a daily basis. Tom talked in parentheses. Each thought as spoken gave rise to a new one that was explored and expanded until the compiled tangents collapsed back onto the main theme. Interactions with Tom were full of conversational oxbows in the sweetwaters of discourse.

Bob was slightly different. I suspect that when he was a child, he must have been scolded for not chewing his food properly because no thought was allowed to be expressed aloud until it had been properly chewed. Watching Bob express a complex idea is like watching a dog worry a piece of gristle. Whereas Tom specialized in thinking aloud, Bob's forte was speaking internally. In the early days this was a cause of great difficulty. When a student went to seek advice from Bob there was inevitably a long pause between the time the question was asked and the response given. (The sweet water of advice came as rather slow drips!) The biggest mistake, however was to think that perhaps the question was not stated clearly, and in an attempt to open the tap to a somewhat steadier stream, the student might undertake to prime the pump with an explanation or another question. This was what Ledyard Tucker used to refer to as "A blunder." A poorly timed clarification would cause the internal speech processes to pause, change directions and reconsider everything that had been done in light of the new information. I know of students who asked Bob for the time of day and later modified the request to distinguish between Central Daylight time and Greenwich Mean Time. They may still be waiting. To say that Bob is reflective is like saying Bill Gates is well off.

During the mid 60s, (the years in which this retrospective focuses 1964-67) in their approaches to evaluation, Tom and Bob also complimented each other. Whereas Tom had a relatively consistent (perhaps constant) philosophy, that, looking back, I would characterize as personalized or customized with a fairly heavy emphasis on assessment. Bob changed quickly. In 1965 we had a series of informal Tuesday morning meetings (Tom, Bob and the 4 or 5 students). To give you an idea of Bob's thoughts on the matter, here is what he said on March 2, 1965 (amended March 3):

> Educational evaluation may be defined as the total description of input and outcome of educational programs. Description of input would include description of the physical plant and equipment, the student and staff personnel, and the schedule and technique of instruction. Description of outcome would include the description of change in behavior, skill, ability, attitude and aspiration among students, teachers and all staff personnel. Where relevant, changes in parents, patrons, and citizenry would be included. Outcomes would not be limited to the implicit and explicit objectives of the program, though the description might be organized around them. Where the expectations of educators and others are relevant these also will be described.
>
> The evaluation of an educational program, if defined as descriptions of input and outcome, cannot be a description of relationships. It is instead a description of coincidence. Relationship requires replication. No relationship is indicated in a single instance regardless of the complexity of input and outcome.
>
> From successive evaluations will come generalizations i.e., descriptions of relationship between input and outcome. Relationships can be described as contingencies with probabilities for each outcome following any input. This process of generalization we might call extrapognosis, and the relationship an extrapolandum. Measurement differs from evaluation in this context in that for measurement the dimensions of variables are given. A necessary part of evaluation is the selection of dimensions or variables or characteristics which are needed in order to effect a complete

> description of the educational programs. The criterion for inclusion of a dimension is variance rather than utility; a dimension is included if it is a basis for indicating the worth of a program. Judgment as to adequacy or merit or worth of a program is not here defined as a component of evaluation. This process is one of scaling the discrepancies between needs or wishes of a community and the outcomes of its educational programs. This scaling procedure depends on other information (other evaluation) than evaluation of education programs, namely the description of needs and wishes. Educational decision making, then, is still another step beyond merit judgment. With judgments made and estimates made of future costs (inputs) and preferences assigned to alternate outcomes, decisions can be made to initiate and direct subsequent educational programs.

If this is early sweet water, we can see that a lot of sediment had to be moved.

As the stream cut deeper, The Countenance paper emerged (thereby giving Mrs. Hull and Mr. Tykociner their brief moments of evaluative fame). Judgments and standards are now seen as important components.

Finally, at the delta we find The Case Study. Judgments are now the central features of evaluation method. A reliable informant has told me of a recent evaluation (and this is beginning to sound more like Whitewater than Sweetwater). Bob flew into town and "cased" the project. One night a hand appeared and wrote on the director's wall:

> MENE, MENE TEKEL UPHARSIN
>
> [For those whose education was entirely secular, the translation according to the Gideon Bible in my hotel room, Daniel 5:25-28 is: God hath numbered thy kingdom and finished it. Thou art weighed in the balance and found wanting. Thy kingdom is divided and given to the constructivists and the critical theorists.]

For Bob, the countenance of evaluation is a changing, growing thing. This gave lots of headaches to his students during the time that the growth curve was at its steepest. Like Terry Denny mentioned last night, often we could not understand the latest version and when we finally did, Bob had moved on. What uncertainty! What an exciting time that was!

While Tom's roots were Muncie, Indiana; Bob's were Adams, Nebraska. I suppose that coming from the farwest of the Midwest, Bob had a greater need to seek his roots. Many of you know that one of Bob's hobbies has been to trace the family tree. Whenever he found himself in a strange city with an hour or two to spare, Bob would look up family names in the local phone book and then call them to see if they were related. [That Bob is a fun guy to travel with.] Using this fundamental case by case incremental approach, Bob was able to trace the family tree to 6 begats from Moses. Then he grew a beard. During his retirement, Bob will be writing about the accomplishments of his ancestors. The first one is to be called, "Quieting Reform: Ten Commandments as Suggestions." As Mike Atkin noted last night, those of you who have received Christmas cards from Bob over the years will know, the family tree has become so extensive that it can now be shown that all evaluators are blood relatives. This is what is commonly known as the problem of relativism in responsive evaluation.

CIRCE was not just sweet water; it was also an extended family. Picnics in the park were an important part of the binding process as were the lists of people that appeared from time to time in the mail. Keeping up to date on the family is an important component of the CIRCE culture.

So what did I take from those three glorious years? Well, Tom was right, there weren't any cookbooks, all evaluations are different, they are explorations both responsive and responsible. And Bob was right--growth and development are essential for evaluators. Don't be afraid to change your mind. Think about what you do. And sometimes it's a good idea to let silence pass unchallenged. As students, education (never training), experience and reflection were the significant themes. We were exposed to some of the best thinkers in evaluation, our ideas were always treated with respect, and we were included in the activities of the Center. We did become part of a family. It was an example that I have tried to recreate for my students during the 30 years that followed.

For over 30 years, CIRCE has been an important contributor to evaluation thought in America and beyond. Although there have been many people who have been part of the CIRCE team, Tom and Bob made it happen. The idea for a

Center for Instructional Research and Curriculum Evaluation was proposed for funding in 1963, at a time when the Department of Psychology and the College of Education of the University of Illinois were enjoying international prominence as significant places for research and development in psychometrics, student learning, and curriculum development. It was natural therefore to put forward a proposal for national funding for a research center. In retrospect it may have been a good thing that the proposal was not successful. CIRCE had the freedom to innovate in a way that may have not been possible otherwise. The various projects that were undertaken stimulated the interests and skills of the faculty and students to produce a legacy of evaluation thought that is unsurpassed in over three decades.

CIRCE--it is sweet water. Thank you Tom wherever you are. Thank you Robert. Ice cream with grape nuts is still a treat.

Lou Smith

Les McLean

David Hamilton

Two Measurement Guys gone Wrong:
or
Fumbling and Stumbling Toward a Paradigm

Louis M. Smith
Washington University

After all of this superficial hilarity of the last couple of days, now it is my duty, yes, my duty to report to you a sad, sad tale, yes even a tragic story of two social scientists, yes real scientists--both included in and attested to by the Jacques Cattell Press' 1962 10th edition of *American Men of Science.* Real men! With bright promising futures! What could be brighter and more promising than their early major publications--with titles like:

Learning Parameters, Aptitudes, and Achievements
and
The Concurrent Validity of Six Personality and Adjustment Tests for Children.

Doesn't that sound like real science? And published in such prestigious journals as

Psychometric Monographs
and
Psychological Monographs.

Who could ask for anything more. What brilliant hopes and possibilities!!!

And where did they end up three or four decades later? Let me tell you. They ended up in a big, fat, wordy volume with the title *Handbook of Qualitative Research.* A handbook yet! And edited by a sociologist--my god--and an educational administrator type--oh, my god, my god!!! And you might ask, what did these two once promising measurement guys author in this fat wordy volume? You won't believe it.

Case Studies
and
Biographical Methods

Think of that! What happened to old E. L.'s dictum:

> If it exists, it exists in some amount and can be measured!

Or the then current adaptation in our guys' graduate years--posted on the walls of research centers. Listen to this.

> If it can't be punched on an IBM card, it's literature.
> And to hell with it.

With rallying cries like those, now long gone, you can see it is a sad, sad, tragic tale I am reporting to you.

Now we must ask, who are the villains of this major tragedy? Well, as you might suspect, they are everywhere. And some of you will know them, but better for me to leave them unnamed. Blasphemous words and name calling are not my style--even with villains. But do let me indicate a couple of the villainous organizations.

The first, you wouldn't even guess--AERA. Recognize that one? That's where real research use to be reported on and discussed. Well, some years ago they created something they called *AERA Monograph Series on Curriculum Evaluation*--seven volumes in all. Rand McNally, the company that publishes those first rate maps and globes and that kind of good stuff, was enticed to publish the materials and became implicated as well. And you might guess who got shanghaied to be the general editor of the series--yes, one of our measurement guys. Villains were lurking everywhere. And, yes, you could guess by now that in the end of that series, in their seventh volume, they snagged both of our guys. That volume carried the longer title *Four Evaluation Examples: Anthropological, Economic, Narrative, and Portrayal*. Can you imagine a monograph like that? One of our measurement guys wrote the first essay:

> *Education, Technology, and the Rural Highlands*

and the other wrote the concluding essay:

> *An Evaluation of TCITY,*
> *The Twin City Institute of for Talented youth, 1971*

But the villainy didn't end there. Our second guy was seduced into beginning his essay with a quote from our first guy. You will remember the opening words

> A *full* evaluation results in a story . . .

Tears come into my eyes, the infamy would not end. One guy implicating the other. It is beyond belief and imagination.

But there were other villains. A couple of them--real measurement men--from old Leland's farm on the west coast - had a six week summer conference in Palo Alto. Not a day or two, or even a week--but a month and a half, six long weeks for indoctrination. And what a place for sin and corruption--you all know about California. One of our guys was in a group on individual differences and learning--they still had hopes of saving him. The other was in a group on social factors in learning--chaired by a social psychologist--and we all know about them--a clear acknowledgment that our second guy was beyond saving, already lost. Give him a little shove along the way! That's part of the nature of evil doers. Oh the havoc you can wreak in six weeks!

But as you might guess, villainy is world wide. So we come to our third set of evil doers--the second half of Oxbridge, that place on the Cam River, in England, no less. From both Scotland and England came the invitation, cloaked with the good words of "alternative methods of curriculum evaluation," and "explore guidelines for future developments in the field," and money from the Nuffield Foundation to pay our way (sin was everywhere), living in Churchill College, oh what would poor old Winnie would have said about that, and evil places in the basement with mysterious names like "the buttery," who had ever heard of that, and serving poison, a dark brew called Guinness. Small wonder that our good guys were lost before they started. And presentations on topics like "illuminative evaluation," sort of talky talk about evaluation rather than real evaluation with real data. And at the end, the evil doers created a "manifesto," and you know the kind of people who do manifestos, a full two pages of rallying cries about "over attention to psychometrically measurable changes in student behavior," and with words like "responsive," "relevant," and "accessible language." Who would have thought our two measurement guys who had started off in the revered psychometric monographs and psychological monographs would have signed off on all this. A real international conspiracy was afoot and our guys were losing, losing badly.

There seemed to be one last chance for salvation. NSF, the National Science Foundation, came into the picture. Now here was a real organization safe from all the evil doers--so one might have thought. But no, one of our good guys was asked to be project director. And the title of the endeavor--*Case Studies in Science Education.* He should of known. And maybe he had a hint for he turned to a number of old friends for help and support, and you can guess already, his old measurement buddy from psychological monograph fame and promise was one of these. Maybe they could pull it out. But no, it was impossible. Our second guy, looked for solutions everywhere, even in a beer hall in Munich where he asked another of our guy's friends and collaborators about independence of the effort and can you trust all the people from NSF on down--or was it up? Our second guy went down in flames, even though he tried to cloak the name with the "Alte School District," an abbreviation of the German word for older suburb, Alte Vorstadt. But case studies they were and remain. And NSF now a part of the conspiracy. It's too much! Overwhelming!

And that takes us back to our handbook villains, that sociologist and school administrator type, with their fat wordy tome of a handbook, published by a company with the seductive title of SAGE. Can you imagine? And what, in the old days, the word sage really meant? But focus on the last episode in that part of the tale. What do villains do with a book that's too big?--for sure you don't reduce it to manageable numbers instead of words upon words--as a real scientist would do. Rather, they cut it into three parts and gave one part a fancy title, *Strategies of Inquiry.* Now doesn't that have a nice sound. And that's where "case study" and "biographical methods" now rest. Pax vobiscum, as my old college roommate would say.

Finally, after all of this sadness, I don't want you to live with total pessimism. Rather, think of that troubadour of the sixties, Arlo Guthrie and his infamous Alice's Restaurant

Massacre and recall his famous lines when tragedies become myths:

> if there is only one of you they will think you are "really sick" and leave you alone
> if there are two of you they will think you're faggots and have nothing to do with either of you
> if there are three of you, you have an organization
> But if there are fifty or more of you, yes, fifty or more you have a movement

And my friends that is what we have now, a fumbling and stumbling movement. And, dear friends if you happen to be a part of a tragedy as happened to our two measurement guys gone wrong, make sure you have a guy like Bob Stake to "walk right in, edge around the back, just a half mile from the railroad track" to travel with. And can it be otherwise, that all of our heroes and all of our villains now say "Thanks Bob."

35 Years Goes Fast When You're Having Fun

Les McLean
OISE/UT (ret.)

Concerned about the intellectual rigour of the presentations at this Symposium, I was determined to raise the level by introducing some scientific content. Something is needed to stiffen the cognitive spine of all these case studiers, relativists and responsive evaluators. There was no need to leave social science, because there are new findings about our perception of time.

Science has recently given us a clear explanation as to why time seems to pass faster as we grow older, and I have been able to extend this finding to explain the link to having fun. Some background is needed, however, in order to share my explanation with you. I first met Bob Stake when he passed through Madison, Wisconsin in 1961. He was taking a promising young student on a tour of graduate schools, and we were deemed worthy of a visit. The student, Gene V Glass, liked what he saw enough to come to Wisconsin, and I have been grateful to Bob Stake ever since. We were in awe of this Robert Stake because of his thesis on complex analyses of cognitive processes[1] (fitting a modification of a rational hyperbola derived by Thurstone), but I remember having fun during his visit and thinking how quickly the time flew by. Looking back, I see this was the beginning of an insight that would take 35 years to fructify. (The present learned paper was begun in 1997.)

The anatomical and biochemical bases of our time perception. The computers on which we are increasingly dependent all have clocks in them, and so must our brain. The clock in our brain apparently keeps time in minutes, or at best seconds (nothing like milliseconds). We are not talking about raw reaction time here: the shortest interval being that from the time the traffic light turns green and the person behind you honks his horn. (It is always "his," of course.) Our concern is with the difference between "My, I thought that bore would never stop talking" and "Where did the time go?" We constantly monitor the progress of external events and respond according to our perceptions--sometimes with anger, sometimes with disbelief. What the scientists have found, and this will bring me to my point (you will be relieved to hear), is the chemical that plays a key role in controlling our mental clocks. That chemical is dopamine. "When the brain notices something new or rewarding, dopamine is released into the spiny neurons, which become excited and begin to integrate time signals. A cluster of neurons in the midbrain collects time signals from all over the human brain and co-ordinates those that occur at the same time and involve singular events or perceptions. Add dopamine and the clock runs faster; take it away and the clock slows down."[2] When our clock slows down, we get our time estimates wrong; nearly 5 minutes goes by and the old folks who are short on dopamine think it was just 3 minutes.[3] The dopamine process is also associated with our feelings of elation and pleasure.

---

[1] See the Abstract at the Toronto site of the Worldwide Stake Celebration Web:
http://www.oise.utoronto.ca/~lmclean/stake/rsindex.cgi

Bob Stake's Influence

Think about it: "When the brain notices something new or rewarding . . ." This is an experience we have repeatedly when Stake is around--he's constantly presenting us with something new and rewarding. We just begin to understand models of cognitive processes when he leaps out of the telephone booth with Antecedents, Transactions and Outcomes. We settle down to cope with the Description Matrix and the Judgement Matrix (lovely rectangles!) and what does he hit us with? A circle! But more important, it can be seen as a clock, the responsive clock, one that does not just go around but that jumps forward and backward.[4] Our spiny neurons are in a perpetual state of excitement and they integrate time signals to beat the band (if you'll pardon the expression). Stake is a dopamine stimulant! Time does fly by, but we know it and we enjoy every moment, because, as you all know, an increase in the presence of dopamine in the user's brain is what triggers a cocaine high.[5] How did Bob get to be this way?

---

[2] Meck, 1996, *J. Exp. Psychol: Animal Behavior Processes, 9,* 171-201

[3] Mangan, 1996, *New Scientist*, Nov. 23.

[4] Stake, Robert (1975) To evaluate an Arts program. Chapter 2 in Robert Stake (Ed.) *Evaluating the Arts in Education: a responsive approach.* Columbus, Ohio: Charles E. Merrill. Pp. 13-32.

As a doctoral student Stake was a Princeton-ETS Psychometric Fellow--crème de la crème of the measurement society of its day. In the early 60s, Bob created a test designed to predict the competence with which graduate students will handle the quantitative aspects of research and advanced study--The QED (*Quantitative Evaluative Device*). It had "parallel" forms, percentile ranks, the works–everything but Rasch scaling. Oh, Stake believed in measurement. But as we all know, he turned away from the measurement path—the preoccupation with quantifiable commonalities—in favour of responsive evaluation and case studies. He questioned the validity of tests of teacher competence in a debate with Jim Popham.[6] My listeners here will be familiar with Bob's view of cases: "We are interested in them for both their uniqueness and commonality. We seek to understand them. We would like to hear their stories."[7] Hmmmm–he hears voices; and he wants us all to hear them--to seek to understand them--without formulas. Must we be drugged?

There are other voices. The sociobiologist Edward O. Wilson, in *Consilience: The Unity of Knowledge*,[8] argues, "The central idea of the consilience world view is that all tangible phenomena, from the birth of stars to the workings of social institutions, are based on material processes that are ultimately reductible . . . to the laws of physics." Nonsense! A dissenting reviewer captured my view (and, in spite of his doctoral thesis, perhaps Bob Stake's): "Measurement always strips away the creative and the unspeakable."[9] We can't hear stories that are ineffable, and following Bob Stake's lead we wish both to be creative and to find and appreciate creativity wherever we can. Following that lead has certainly done a lot to keep my dopamine flowing; 35 years have passed, but it has been--and will continue to be--fun. We're not about to give it up, eh Bob? Let's give the last words to Tennyson:[10]

---

[5] *Nature*, April 24, 1997, vol. 386, p. 827

[6] The text of Bob's presentation at the debate is also at the Toronto website.

[7] *The Art of Case Study Research*, p. 1, emphasis added.

[8] Knopf, 1998

[9] Richard Lubbock in the *Toronto Globe & Mail*, April 18, 1998.

[10] "Ulysses," second and final stanzas.

How dull it is to pause, to make an end,
To rust unburnish'd, not to shine in use!

……

Tho' much is taken, much abides; and tho'
We are not now that strength which in old days
Moved earth and heaven, that which we are we are;
One equal temper of heroic hearts,
Made weak by time and fate, but strong in will
To strive, to seek, to find, and not to yield.

Soy(a) Bean Futures near the Arctic Circle (or How Green was Bob's Volvo?)

David Hamilton
Institutionen för pedagogik
Umeå universitet[1]

I had prepared this presentation from notes I made on my way to Champaign for the Stake Symposium. I was invited to contribute to a symposium on "Assessing, evaluating, knowing." I knew it would be an occasion for celebration, retrospection and introspection. I had no idea what I would follow. I did not know where to start--except from somewhere far out. At the event, I decided to linked my time at CIRCE in 1976--redolent of soy beans and Bob's green Volvo--with my current position at Umeå University in Sweden (where Bob had trialled *The Art of Case Study*). Hence my title.

Ours was the second symposium on Saturday. Key words in the previous symposium included interpretative turn, consequential validity, judgement, democracy. I wove them into my presentation, even though I felt that the key issue raised by the symposium was not so much the demonstration of democracy as the resolution of the *problems* of democracy.

I started with two "items." That 11 European countries had agreed a common currency the previous Sunday, and that the Chrysler Motor Company had announced that it was to be taken over by Daimler Benz. I then posed the question: What complex processes of assessment, evaluation and knowing had gone into these decisions? I generalised--naturalistically, of course--to the soy(a) bean question: to the essentially qualitative assessing question (of what kind of bean are we talkling about?), to the subsequent evaluation question.

---

[1] 901 87 Umeå
tel: +46 (0) 90 786 6564
fax: +46 (0) 90 786 7800

Theresa Souchet, Marya Burke,
Chris Migotsky, Rita Davis,
Edith Cisneros-Cohernour,
Mindy Miron Basi

Iduina Chaves

High Expectations at CIRCE: Bob as Mentor

Theresa Souchet, Marya Burke, Christopher Migotsky,
Rita Davis, Edith Cisneros-Chohernour
Mindy Miron Basi

Poem:
In the twilight of a long teaching career
is when I get the pleasure.
A pristine reputation for working with students
he has not.
I am trepidacious.
A dignified scholarly demeanor keeps me distant,
despite offerings of mid-day hot chocolate.
Always busy, often swamped.
Much vying for his mind.
When can I ask?
Meekly,
Do you think you could help me . . ."

Sure. Let's talk now.
(Suddenly, it occurs to me how often I've heard
those words.)

One Liner:

During my time at CIRCE, my fellow grad students and I lived in terror of "disappointing Bob." It was easy to do. Only exemplary work won his approval. *Linda Mabry*

Narrative 1:

"Bob, can you clue me in on this concept, I am not sure I get it." Forty-five minutes later I leave with thoughts tingling in my head. At the computer I let ideas form, reasoning "if I write, it will come." With a vague uneasiness, I turn a draft over to Bob. A few days later I retrieve it, heavy with ink in his rocky calligraphy. "No that's not it, keep trying." I sigh and write myself into his appointment book. After another long discussion, I try again. A cloud of understanding floats just out of range. Squinting, I make a little bit more out. After several weeks and too many drafts, Bob pops his head out of the office. He notices my strained expression and peers at the screen, the source of my discomfort. "Are you still working on that?"

One Liner:

When I finished a review for a paper Bob asked: "How long did this take?" "About 8 hours," I said. He chastised: "Some things deserve 8 hours. Most don't." He walked away. *Deborah Trumbull*

Narrative 2:

Bob, give a straight answer?! That would take the fun out of it. He prefers to leave us in constant doubt. He always gives us feedback, feedback which is often hard to take. I remember taking his Case Study class. I had a lot of trouble trying to write the issues for my study. I kept writing research questions--breaching one of his pet concerns. He repeated again and again: "Issues come from the case, you need to draw on both reasoning and your intuitions." On one occasion, he told me not to worry, and added that I will know them when I see them. I tried again. This time he said: "You are in your third year in your program and you are an education specialist, you should know this by now. You are just not trying." His next round of criticism was less harsh and more motivating, "You can do this. You know you can. Go to the library and read the literature. You need to remember why you care so much about this topic."

One Liner:

On asking Bob if he had read my paper, he replied "Oh yes. It was like an in-grown toenail." *Stephen Kemmis*

Narrative 3:

"Perfect. A piece of art work," this was on my first evaluation! I am not one to emote, but I felt like crying or cheering or jumping up and down! I had to share the good news. "Mindy, look at this!" Her surprise confirmed my elation. Of course there were some minor corrections. I zipped through them and resubmitted the report, looking forward to final "certification." When my "work of art" was returned, there were quite a few splotches of red paint on it. It seemed some areas needed substantial revisions. It may have been a piece of art, but in some important ways, it had failed to meet this critic's expectations.

One Liner:

Stake was frustrated with me. He wanted me to settle on a case. I worried about becoming a basket case! Ernie House helped by telling me that NEARLY becoming a basket case was a key aim of the game, but that ACTUALLY becoming a basket case spoiled everything for Bob. *Robin McTaggart*

Narrative 4:

His exactitude and attention to detail was not just limited to the content of a case or the analysis of a case. It also included the writing of a case. I learned from Bob that it was not just what you wrote, but how you wrote it. His edits were always correct; his analyses of the weaknesses of the piece were always on target. Those red marks brooked no space for objections--he was right. He taught me that sentence structure, down to the use of a single word, makes the difference between clarity and confusion.

One Liner:

While taking his class I was confused. I thought I was doing great. But I didn't really have a clue. It was actually kind of refreshing. *Anonymous*

Narrative 5:

Bob's frown is second only to his smile in communicating approval or disapproval and it is these subtleties that are part of the feedback I value. During a client briefing this past year, Bob had endured my nerves and fretting over my role at the meeting for 24 hours. When my turn finally came to speak, I said to the clients gathered around the conference table, "Yes, our panel found the real letters more responsive than the performance task letters, but that may have been due to. . ." Out of the corner of my eye, I saw Bob's smiling nod--a silent, but powerful message.

One Liner:

Paradoxically, "If ever Bob put you on the spot, or asked you to do better than your best, maybe you too, have the misfortune to be his friend. *Stephen Kemmis*

A Brazilian's Stakean Journey

Iduina Mont'Alverne Chaves,
Federal Fluminense University

> At 8:30 a.m. on Thursday morning, Adam shows up at the cafeteria door. Breakfast is being served but Adam doesn't go in. The woman giving out meal chits has her hands on him, seems to be sparring with him, verbally. And then he disappears. Adam is one of five siblings, all arrive at school in the morning with less than usual attention. Short, with a beautifully sculpted head and Gerri-curl, solid body baggy black sweats and sneakers, and full of energy, Adam is a person of notice.
>
> . . . It's Mr. Garson's fifth-sixth grade room. Garson notices Adam, has a few quiet words with him before a paternal shove toward the room.
>
> . . . It's a typical elementary school room with full windows on one side, blackboards across the front, homemade and purchased posters almost everywhere (Stake, 1995).

The quotation above include pieces selected from the "Shadow Study of a Sixth Grader" drawn from Stake's "Harper School" case study report. It is a very touching, Bob's influential attempt to vicarious experiential knowing. To Stake, case study is compatible with experiential knowing and enhances opportunity to increase understanding of teaching through disciplined attention to detail, vicarious experience, multiples realities, context. To him:

> vicarious experience is telling, and so we tell it. Vignettes sink into our consciousness at a level deeper than linguistic coding. Scenes and nuances become background, prior knowledge, against which future perceptions will be framed. (1994, p. 34).

Much has to be learned from Bob Stake's thinking, as a researcher, as a teacher and most especially, as a human being.

In this paper I want to consider some of Stake's discussion on the nature of qualitative research stressing his observations on *constructivism* and *interpretation*. I also intend to discuss how his ideas nourish research on Brazilian education.

In 1995 I joined the Center for Instructional Research and Curriculum Evaluation (CIRCE) team during a one year scholarship under Bob Stake's orientation. I participated in an evaluation study of the Teachers Academy for Mathematics and Science, a study of the quality of staff development activities. At the same time I was attending classes on "Case Study Research Methods," "The Theories of Educational Evaluation," and "Qualitative Data Analysis" led by Dr. Robert Stake.

It was Winter-time. I attended a meeting at the Teachers Academy in Chicago. Full of energy I arrived at the windy city. I received a cheerful welcome by Academy's people. I had my eyes and ears open to grasp every movement and to register each activity. I maintained informal talking with the principals and school teachers around tables of a classroom. Later I joined the group for a friendly chat during Coffee and Lunch break. It was a pleasant day. And a good opportunity to learn about schools and teaching in Chicago. The way back home was a special moment to ruminate and organize my thoughts on data gathered.

Back in Urbana I prepared a formal report to Bob, the Director of the CIRCE research team. It provided a chronological and detailed description of everything observed, and was considered a good report. But Bob asked for a new exercise: "try to choose and write up a particular event, describe an episode that could be significant to illustrate your reflection on the meeting and enhance the most important findings. And add your own reflections about it."

I left the room full of anxiety but ready to face the challenge: to write a report following Bob's recommendation. Issues, vignettes . . . Stakean's lessons, so peculiar but still needing a greater awareness about this new reality and an expression in my practice. I revised my classnotes, talked to more experienced peers, read alot of reports on qualitative research as Bob had advised me. I understood that I had to be more intuitive than rational....

> I paraphrased Bob's words; and repeated to myself again and again: "I have to be a provocateur of understanding, I have to portray the common in problematic ways." The act of creation, of construction (knowledge) is somewhat an instigated moment, but it is suffering too. Especially for me, a newcomer to this new paradigm on qualitative research, to this new look to understand and interpret reality.
>
> The report was writen in narrative form. It pleased Bob.
>
> This experience opened a window to me. I had the opportunity to read, to pay attention and to learn about how to write up reports on qualitative research.

First it was important to get the distinction between qualitative and quantitative research according to Stake. Three major differences deserve attention:

> (1) the distinction between explanation and understanding as the purpose of inquiry; (2) the distinction between a personal and impersonal role for the researcher, and (3) a distinction between knowledge discovered and knowledge constructed (Stake, 1995, p. 37).

Quantitative research demands explanation and control whereas qualitative research presses for personally understanding the complex interrelationships among different realities. In other words, explanation is attached more to propositional knowledge, while understanding is linked to tacit[1] knowledge. To sharpen the search for explanation, Stake (1995) says that:

> quantitative researchers perceive what is happening in terms of descriptive variables, represent happenings with scales and measurement (i.e. numbers). To sharpen the search for understanding, qualitative researchers perceive what is happening in key episodes or testimonies, represent happenings with their own direct interpretation and stories (i.e. narratives). Qualitative research uses these

---

[1] Tacit knowledge understood by Stake (1978), includes a multitude of unexpressible associations which give rise to new meanings, new ideas, and new applications of the old (p.6).

> narratives to optimize the opportunity of the reader to gain an experiential understanding of the case (p. 40).

To Stake, the centrality of interpretation[2] is the primary characteristic of qualitative research.

Von Wright (quoted in Stake, 1995) illustrates the distinction between explanation and understanding:

> Pratically every explanation, be it causal or teleological or of some other kind, can be said to further our understanding of things. But "understanding" also has a psychological ring which "explanation" has not. This psychological feature was emphasized by several of the nineteenth-century antipositivist methodologists, perhaps most forcefully by Simmel who thought that understanding as a method characteristic of the humanities is a form of empathy or re-creation in the mind of the scholar of the mental atmosphere, the thoughts and feelings and motivations of the objects of study. . . . Understanding is also connected with intentionality in a way that the explanation is not. One understands the aims and purposes of an agent, the meaning of a sign or symbol. And the significance of a social institution or religious rite. This intentionalistic dimension of understanding has come to play a prominent role in more recent methodological discussion (p. 36).

A summary of the characteristics of qualitative studies devised by Stake (1995) is suggestive. He speaks of qualitative inquiries as *holistic, empirical, interpretive and empathic*. The *holistic* characteristic reflects its contextuality, that it is a case (a bounded system) situated, that it resists reductionism and elementarism, and it is relatively non-comparative, seeking to understand the object itself more than to understand how it differs from others. Qualitative inquiry is *empirical* because it is field oriented, its emphasis on observables, including the observation by informants. It strives to be naturalistic, non-interventionistic, with a preference for natural language description. It is *interpretive* because its researchers rely more on intuition, with many important criteria not specified. Its on-site observers work to keep their attention free to recognize problem-relevant events. It is attuned to the fact that research is a researcher-subject interaction. The *empathic characteristic* attends to actor intentionality, it seeks the actor's own frames of reference, value commitments. Although planned, its design is emergent, responsive. Its issues are emic issues, progressively focused and its reporting provides vicarious experience (p.47-48).

---

[2] Stake makes references to the work of Egon Guba and Yvonna Lincoln (1982); Elliot Eisner and Alan Peshkin (1990); Henrik von Wright (1971); and Frederick Erickson (1986).

In his book *The Art of Case Study Research* (1995), Bob Stake provides a significant contribution to research, epistemology, and practice. It's a pleasant trip into the methodological field of qualitative inquiry. It is the image and resemblance of Bob in the classroom. And in the campus classroom situation, Bob Stake is a wise researcher who teaches. For him *teaching* is one of the major roles of the researcher.[3] As the intention of research is "to inform, to sophisticate, to assist the increase of competence and maturity, to socialize, and to liberate . . . these are also responsibilities of the teacher." He adds that:

> teaching is not just delivering information; more, it is the arrangement of opportunities for learners to follow a natural human inclination to become educated. Providing information, arranging access to information regularly, is a major part of teaching, but two prior considerations are the selection of information and/or experiences needed and the recognition of conditions that will facilitate learning for learners individually and collectively.

As researcher, the teacher is an advocate, he "is the exemplar of the way to see, the persuader of a road to follow." Stake claims also that the more the teacher knows the individual faces and their minds, the better would be the teaching. It is also true for researchers who try to teach their readers. Considering this, Bob Stake poses some elucidative questions:

---

[3] To Stake (1995), the case researcher plays different roles that include teacher, participant observer, interviewer, reader, storyteller, advocate, artist, counselor, evaluator, consultant, and others. Each researcher makes continuous decisions about how much emphasis to give each role. (p.91)

> How familiar are the words, how similar are the experiences, how attractive are the vignettes and assertions that populate the report? Most prospective readers are not close at hand. It is important to create imaginary readers to worry about their needs. What to them is comprehensible? What will be remembered? What will be contested?

He recommends the use of ordinary language and narratives to describe the case and the opportunity for readers to make their own interpretations along with the author's. In Bob's own words:

> with effective description of persons, places, and events, the research provides a vicarious experience which readers can attach to other knowledge about teachers and teaching. If the new knowledge is persuasive, the old is amended, revised or, on some occasions, thrown out. Theorists, researchers, teacher educators, and teachers–we all come to know in this way (Stake, 1994 b, p. 34).

We can feel here a great consideration for the readers *with the purview of good teaching*. In his own words:

> to assist the reader in making generalizations, case researchers need to provide opportunity for vicarious experience. Our accounts need to be personal, describing the things of our sensory experiences, not failing to attend to the matters that personal curiosity dictates. A narrative account, a story, a chronological presentation, personalistic description, emphasis on time and place provide rich ingredients for vicarious experience (Stake, 1995, pp.127-29).

Bob Stake, as a constructivist, is coherent in what he says and what he does, as a researcher as well as a teacher. What one can learn from him is substantial:

> Infants, children, and adults construct their understandings from experience and from being told what the world is, not by discovering it whirling there untouched by experience. In schools, they study science, memorizing the answers and doing experiments. What they know of reality is only what they have verified outside their experience. . . . Human construction of knowledge appears to begin with

> sensory experience of external stimuli. Even in the beginning, these sensations are immediately given personal meaning. Although originating in outside action, only the inside interpretation is known. As far as we can tell, nothing about the stimulus is registered in awareness and memory other than our interpretations of it. No aspects of knowledge are purely of the external world, devoid of human construction. In our minds, new perceptions of stimulation mix with old. . . . Although the reality we seek is of our own making, it is a collective making (Stake, 1995, pp.100-102).

The literature on constructivism is extensive, spread out everywhere. But it is worth saying that one can feel it misunderstood. Bob Stake argued that most of time we think constructivism it our choice, as researchers, to follow one methodology or another, one epistemology or another.

> But what we choose to believe in, as evidence, is more determined than volitional, more intuitive than rational. As searchers, we find the deeper question of constructivism: "what constitutes evidence? Why is one image better testament then another? . . . Confirmation is not the aim of constructivist research. Composition is not the aim of constructivist research . . . constructivist fieldwork seeks unrealized problems among familiar settings. From performance, from interpretation, awareness of the multiplicity of realities is sharpened (Stake, 1994, p.42).

How Bob Stake's ideas contributed to qualitative research in education in Brazil

I was one of Bob Stake's students and he was my supervisor during my graduate studies in Illinois. His influence towards the background I brought from Brazil was incommensurable.

As a teacher, his peculiar way of leading the teaching/learning process is commendable. His action in the classroom portrays the often challenged interrelation between theory and practice, teaching and research and between content and methodology. I want to explicate some of the contributions to our educational reality as well as to my professional development, as a teacher and as a researcher.

Brazilian Universities present, like those elsewhere, a dichotomy between research and teaching. They are divided into two different activities performed by faculties. They lack integration and flexibility. Education would benefit if some of Stake's proposals for integrating research and learning in and out of classroom were adopted in the everyday life of the university. This means transforming the research procedures into courses that could incorporate theory and methodology and integrate theoretical approaches with usual practice. Teachers should prepare themselves to reconcile qualitative research with teaching.

Bob Stake spent some months in Brazil, sixteen years ago. I interviewed Dra. Menga Ludke, the teacher who invited him to participate in the "International Evaluation Debate Seminar," organized by the Education Department of the Catholic University of Rio de Janeiro, in 1982. Menga is a highly respected professor and an authority on educational research in Brazil. Now, for the last twenty years, she has been teaching Methodology of Research at PUC, and is currently a member of the National Council of Science and Technology (CNPq).

From the vignette below (excerpt from Menga's interview) one can feel Bob's presence and influence in Brazil.

> . . . Bob Stake arrived in Brazil in the early 80s at a time educational research was in a process of changing. He brought a greatly needed view of qualitative inquiry. Bob Stake's presence in the Seminar brought light to many research and evaluation questions we had been struggling with. For example, at that time, I had a master's degree student who wanted to do research on Literacy. She wanted a way of doing research that allowed the researcher be close to the people involved with the problem, particularly, the teachers and their students. The qualitative approach provided a new way of doing research on the individualistic character of Education. My student decided to start a Case Study, perhaps the first Case Study in the field of Education in Brazil. Bob's arrival was exactly at the right time. It was a great opportunity to discuss issues of the uniqueness of illiteracy in Brazil. Bob embraced these questions and suggested ways

to study them. It was the first of many practical lessons from Bob.

Bob's publications and personal conversation gave us testimony to his own understanding that traditional ways of thinking research would not always fit the needs of education today. He reaffirmed here that he had broken with this research based on standardized testing and inferential statistics and that he looked to ethnographic research and Case Study for data gathering and analysis.

In Brazil, Bob Stake can be considered one of the founders of a new approach. In a book written by Marli André and me, we spell out Bob's undenyable influence.

Holding a Fullbright scholarship from June to August 1984, Bob Stake taught at the Federal University of Espírito Santo in the Program of Post Graduate Education coordinated by Dra. Elizabeth Pinheiro Gama. The visit was also informally hosted by Maria da Penha Tres of the State Department of Education. Penha told me:

> During this visit, Bob Stake taught a special seminar on qualitative research methods for faculty members. He also worked as a consultant in a research project called: "Estudos das Disparidades Educacionais no Espírito Santo." And he participated as a consultant and researcher in our two field studies visiting about 30 rural schools in the Anchieta District, E.S. During this Summer of 1984, Bob was the key speaker at the National Debate Coference in Brasilia sponsored by CNPq. He also participated in a National Debate on research methods sponsored by the National Association of professionals in Educational Administrators ANPAE, Brazília, DF., from July 29 to August 2, 1984.

Bob became a permanent mentor for us in our maturation as researchers. My colleagues and I were fortunate to get to know Bob as a person, and to learn from him how to think about qualitative research, specially the Case Study. His stay in Brazil gave us an opportunity to reflect upon the epistemology and methodological bases for our research and evaluation and even to refocus our graduate program, for it was, at that time, in an accreditation process. We considered it an "unmeasurable" privilege to have Bob and Bernardine with us.

My Ph.D. dissertation dealt with a Teacher Education School. I always thought of writing the thesis in a style that could be understood and appreciated by readers. I did not want to write in an academic and sophisticated style, comprehensible only by a small elite. I learned with Bob Stake to think of the reader as a constructor of knowledge, to write so as to maximize the reader's encounter with the complexity of the case . . . and to tell a few stories or vignettes to illustrate my study. I thank him for making me discover a new way to express my ideas in the dissertation--the narrative style. I felt very comfortable to hear at the occasion of my qualifying exam: "Your work is deep and the reading flows smoothly." As a faculty member of the Faculty of Education at The Federal Fluminense University leading Research Practice and Pedagogy classes, I am trying to follow the knowledge I have constructed with help from my Big Master Bob Stake.

We have to thank God for the existence of a Bob Stake in the world. And more, for the privilege of having known him and being around him to receive his lessons.

References

Stake, R. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2).

Stake, R. (1986). *Quieting reform.* Urbana: University of Illinois Press.

Stake, R. (1994). Case study. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). Thousand Oaks, CA: Sage.

Stake, R. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stake, R. & Kerr, D. (1994). *Rene Magritte, constructivism, and researcher as interpreter.* Paper presented at AERA annual conference.

Stake, R. (1994). *Case study: Composition and performance.* Bulletin of the Council for Research in Music Education, Urbana-Champaign, IL.

Stake, R. (1995). *Lasting innovation in Education: Why is it so difficult?* Ben Bronz Academy's tenth anniversary, Farmington, CT.

Stake, R. (1997). *Stress on cognitive teaching in school systems prioritizing*

Stake, R. (1998). *When policy is merely promotion by what ethic lives an evaluator?* Lecture, University of Tel Aviv.

Stake R. & Mabry, L. (1995). *Case study for a deep understanding of teaching.* Course Document.

Stake, R. & Mabry, L. (1997). *Ethics in program evaluation.* Conference on Evaluation As a Tool in the Development of Social Work Discourse, Stockholm.

Mildred Griggs

Jeff Stake

Clem Adelman

Dan Alpert

Barry MacDonald

Terry Denny

Remarks of Mildred Griggs

Greetings. I am delighted to welcome all of you to the Campus of the University of Illinois to celebrate the career of Professor Robert Earl Stake and indeed it is a career that is worthy of celebration. Professor Stake's tenure at this University started in July 1963 with a letter from the Dean of the College of Education, Alonzo G. Grace to Provost Lyle Lanier in which he lamented the imminent loss of two very distinguished faculty. However, he went on to say that he had identified an outstanding candidate with excellent qualifications to fill the vacancy left by their departure—in the person of Robert E. Stake. We all know that Professor Stake has achieved the status of an intellectual giant, and Dean Grace was clairvoyant enough to see that promise in the young Bob Stake and employed him as a replacement for not one but two highly distinguished professors.

We recognize Professor Stake for his leadership and extensive scholarship in educational evaluation. However, his personal peculiarities and mannerisms, those that you have heard described today really endear him to all of us. Let me cite an example of an inimitable Bob Stake mannerism. Back in 1963 when Dean Grace made an offer to Bob to join the faculty in the College of Education, his response was that the offer is attractive, I will give it serious consideration and give you a decision in a couple of weeks.

Bob, we are extremely happy that you did accept the offer for it saddens us to think what life in our College would have been like without you. You have had tremendous influence on all aspects of the College of Education at the University of Illinois. Your legacy touches all of education. We are in awe of the tremendous respect that you have earned over the years that is in part indicative of the warm, touching comments made by colleagues, former students, family and friends who have traveled across the world to be here for this celebration.

Bob, best wishes in the future as you continue to inspire us to think critically about education issues. We are thankful for privilege of being a part of your wonderful career and for the opportunity to be involved in this great celebration.

Remarks of Jeff Stake

Being one of the true Stakeholders here, I guess it is my job to debunk some of the dogma we have heard so far, to share some of my experience, to make some naturalistic generalizations. But how do I generalize about a man whose image in my head is wearing a T-shirt that says "see how each datum differs"? Perhaps the best I can do is supply a few of the data from which you can do your own vicarious generalization about Robert Stake.

A friend and poker playing buddy of mine took a course in Secondary Education Social Studies with Larry Metcalf. One night, during our poker game, he told me that a question of evaluation had come up in class. After Larry thought about it a bit, he said that he did not know what to say. Then Larry added "We could ask Bob Stake, but we don't have time for the answer."

Imagine the irony I felt when I learned later in life that he promotes "responsive evaluation." RESPONSIVE? It should have been called "responsive, after a pause."

But that generalization--slowness and completeness--like any, is unfair. I remember asking him where I should go for my "senior visit day," the day when high school students go to a college to see if they would like to enroll. I told him others in my class were going to Harvard and MIT, and I think Mike Atkin's son was going out to Stanford. He said only, "Try Chicago."

What does he mean?
What does he mean?
Is Chicago a good school?
Would I enjoy it there?
Since we had never taken a family vacation by airplane, my interpretation was that the train ticket was less than airfare to San Francisco.

But there were other meanings, other Truths.

One was that pecking order and rankings do not matter. Some people think that they have an irrebuttable rejoinder: "Don't they matter when it comes to picking a brain surgeon?"

I will give you Dad's answer:

> Dad was lying, dying, in Carle hospital because they could not seem to understand that his colon was disintegrating. We suggested that we go to Mayo Clinic for help. He objected, saying that this was his town, he had lived here and would die here.

Fortunately, Mom's greater wisdom prevailed on that issue. And because of her, he is here to share in celebrating his contribution to our education.

Remarks of Clem Adelman

Some end of the last century French novelist suggested that in the end we all do for a living what we are second best at. I take this to mean that given a choice of alternatives we proceed down the path which does not entail risk to our deepest aspirations. In Bob's case the evidence of whether his sustained productive, creative work is second choice or first is in no doubt. There is no way to such thoroughness without devotion and risk. But Bob has not become obsessed, he has held his work within a wider set of life interests. I will only comment on his feel for incisive music making, his recognition of the integrated concentration of the expressive musician. Now, I know that Bob refrained from becoming a musician beyond marching band but he knows a lot of songs. Given the archeology of our minds it may be that several of Bob's important emphases, detail of the particular, responsive evaluation, principles rather than standards may have stemmed from his musing on song lyrics. This raises questions about the status of anachronism which we have no time to go into here, so I propose we skip those parts of the argument and get down to the real onions . . . The particular tune will be briefly rendered or rendered briefly. We ask you recall the title and the possible influence of this idea on Bob's thinking.

We are, of course, giving prizes for those who at least recognize the titles. Having discussed the matter with Bob we are offering as first prize one week in Las Vegas, second prize three weeks.

Band plays first four bars of Great blues by Billie Holiday "traveling light"!!

It ain't what you do it's the way that you do it
You or nothing at all
I concentrate on you
I'm beginning to see the light
Let's call the whole thing off
It's easy to remember but it's so hard to forget

Remarks of Dan Alpert

I was not a student of Bob Stake, and I have never claimed expertise in the field of Program Evaluation. My friendship with Bob goes back to the time we first met (in 1968, I believe) before most of his current students were born. And you will note some significant differences between my reactions and the wonderful stories that we just heard from the current graduate students.

To me, Bob Stake personifies an ideal teacher as characterized by Donald Schön. (Schön did not use the term teacher or professor; he preferred the term "learning agent.")

> *The LEARNING AGENT must be willing and able to use himself as an informational instrument within the learning situation. His own abilities to listen rather than to assert, to confront and to tolerate the anxieties of confrontation, to suspend commitment until the last possible moment--all condition his ability to draw information from the situation while it is still in progress.*

I learned a great deal from Bob during many conversations, workshops, informal get-togethers, and other professional interactions. I always found that whenever Bob entered a room, it became a *safer place* for me and others to speak candidly, especially about matters that were controversial or sensitive.

Moshe Feldenkrais spoke to the need for a learning environment:

> *To learn, the environment must be safe and pleasant . . . You must get some enjoyment out of it.*

For me, Bob always contributed new ideas, different interpretations, and interesting perspectives. He could express disagreement without becoming disagreeable. Indeed, it is sometimes hard to tell whether he agrees or disagrees; in either case, he leaves space for people with contrary views. He has remained on friendly terms with colleagues who espouse quite different approaches and has sought to embrace multiple perspectives.

Bob has been a leader in the field of Program Evaluation, and I have appreciated his approach. Some of his case studies read like novels, and I once suggested that one such report should have been marketed *that* way. There may be those who wonder why it is still necessary to define or redefine the *field* after these many years. Why hasn't the task been completed?

Reinhold Niebuhr, the eminent theologian, spoke to this question as follows:

> *Nothing that is worth doing can be done in one lifetime.*

There is every reason for Bob Stake to keep up the work that is epitomized by this symposium. It is eminently "worth doing." His style can be summed up in a quote from Etienne Wenger of the Institute of Research on Learning:

> *A productive life-long learner--a person who can adapt and learn swiftly in new Situations--is one who can transform all situations into learning situations.*

Bob: Keep up the good work in the future; I wish you well!

Remarks of Barry MacDonald

I don't remember what I actually said, because I changed it in response to the mood of the gathering and what had already been said, and shortened it because the audience at that point were showing signs of incipient cramp. What follows therefore is a compendium of resources I had in mind when I stepped up to the podium. Do with it as you will--my memory is worse than Clinton's.

If this event marks the end of Stake's career, then it's not just the postcard industry that will regret it. I didn't know an evaluator could have this many friends. After all, our job is to interfere with people who just want to be left in peace to get on with their work. You don't make many friends that way. And if, like Bob, you are almost invariably right, reasonable and fair, you can't expect forgiveness either (Bob, you can argue with me later about what I mean by "almost"). And if, like Bob, you are obstinate, uncompromising and persistent, as well as being right, then one could be forgiven for being surprised that anyone turned up, other than to make sure.

But most of us present are, in one way or another, indebted to his inspiration, including those dissenters who have been compelled to sharpen their refutations to withstand his critique. In the course of these two days that debt has been fully expressed and I feel no need to add to it, other than to say that his influence has been truly international. If you want a measure of his stature, I offer the following conversation.

When I told one of my students that Stake was retiring, he replied, "Oh good. Does that mean we can do what we like now?" I replied, "Fat chance, it's an American retirement, they just take the day off."

I have another observation on that to make. I have it on good authority (if you'll pardon the contradiction in terms) that Stake is only 64 years old. Well, all I can say to that is that I have known Bob for nearly thirty years and that's the youngest he's ever been.

What has received little mention so far is Bob's sense of humour--very active and mischievous I can tell you.

Geography was never my strong suit and for many years my territorial knowledge of the USA was restricted to New York on one side, California on the other and Chicago somewhere in the middle. Bob took a mean advantage, and it was a long time before I realised that it was extremely unlikely that he had in fact served in the Nebraskan Navy. I remember, too, puzzling about a statement attributed to Eva Baker, to the effect that you can't run Los Angeles as if it was Adams, Nebraska. I puzzled about the distinction until somebody told me that Los Angeles was not a State capitol.

Finally, I would just like to say that the key concept we took from Stake thirty years ago was the notion of evaluation as storytelling. In the UK these are hard times for story tellers. We can still tell them, but our sponsors increasingly insist on a happy ending. With policy in the UK reduced to a choice between blunders, it's a bit like writing the story of the Titanic without mentioning the iceberg. But, as I'm sure Bob will, we'll carry on.

Selected Memories of Robert Stake

Terry Denny

I regard Robert Stake as the leaven of U.S. educational evaluation efforts over the past three decades. He transformed much of *how* we think about evaluation. He gave rise, if you will, to a deeper understanding of the role human judgments play in the process of evaluating educational efforts. He made me stop and question how I conducted my evaluation efforts. Some of you have been privileged to read his insightful, pithy, and provocative one-page pronunciamentos--such as A is A and the Ever Normal Granary. No one did it better. No one does it better.

But that's not the way it began. A recent hernia operation prompted me to recall the second time I met Robert. It was the summer of 1967 . . . he kicked me in the testicles. I saw a thousand points of light. I have been blinking, waiting, ever since to say a few things about him.

The first time we met was kinder. I had just accepted a position with ETS and was supposed to accomplish something with the notion of evaluating school curricula. As soon as I learned how to spell *curricula* I turned my attention to *evaluation*. Everything I knew was based on what Ralph Tyler had written. In my early days when I got troubled I often turned to church, booze or the library. It was in the library that I first learned about Robert Stake.

He had just published his monumental *Countenance* paper. I was enthralled with it. I even thought I understood it. It made some sense out of the scattered efforts I was doing in the name of school evaluation. I had just finished the national evaluation of Catholic schools. It was time to start making sense. So I drove to Urbana to meet him for dinner. It was my intent to enlist him as a consultant for my ETS work that lay ahead.

Although I cannot recall a single concept discussed at that dinner, one day later I resigned my position with ETS, without having worked a day, and signed on to work with another gentleman who was also at the dinner, Kenneth Komoski, the inventor of EPIE.

Let's return to my testicles. Two months later, I am on the so-called payroll of EPIE. I had the privilege of working with Robert and Ken during a summer workshop sponsored by EPIE. Two things I found out in short order that summer: Stake could play table tennis and shoot set shots. He took my lunch money away from me frequently. At no time did his feet leave the ground in either sport.

I learned that he was also not above using the SUNY-Southampton College logo--a windmill--to establish himself as the alpha male in our relationship. One day the man from Adams took off down the hall racing toward me, did a cartwheel on the tiled wind mill logo whilst I stood on same. His foot flew into my crotch and down I went. In my mind that poignant experience was the precursor to his work in responsive evaluation. It is probably why I cannot understand it--and still think that the *Countenance* paper remains the most useful, brilliant and compelling piece that has ever been written about educational evaluation.

Among the many things I have learned from Robert are the following:

- Some redundancies are not necessarily redundant
- Unlike Mies van der Rohe, who thought that less was more, Robert taught us that *less may already be too much*
- The results of massive testing of all children is that we invariably miss every child
- In large scale evaluation efforts he trusted good people to do good things
- Robert didn't believe in original sin, national testing programs, standardized curricula, deans' offices or city hall--but he did believe that Tom Hastings was CEO of CIRCE.
- Probably Robert's largest teaching for me was and is that ***family is not just important --it is central.***

Thirty years ago I used to think that one of the best things about Robert was Sara, Ben, Jeff and Jake--and the incomparable emotional cement of his family, Bernadine.

I used to **think** that was one of Robert's best qualities--now I know it.

Stafford Hood reminded us earlier in the day that President J.Q. Adams made his greatest statement as an 83 year old. Is it too much to expect? Not from Robert Stake. Thank you Robert, for many uplifting experiences.

Elliot Eisner

Bob Stake

It's All About Bob Stake

Elliot Eisner
Stanford University

What is it about Bob that has brought so many so far to celebrate his career, his retirement, and his leadership? Was it Bob or was it frequent flyer miles?

To find out I conducted some research. I did a structured interview with his friends using the Cronbach alpha to determine the inter-item consistency—or is it reliability—of the data provided by the interviewees, which incidentally boasts a ninety-eight point five percent acceptance rate. The one person who declined to be interviewed was a graduate student of Bob's who, when I asked him, couldn't remember the meaning of "naturalistic generalization," and even worse, never heard of "responsive evaluation." Anyway, what could *he* know about Bob?

In any case, I got a sample, though not really a random one, and asked the sample, "What makes Bob so very special;" note that in this open-ended question I was very careful not to bias their response. And because of the twenty bucks I paid all three interviewees to do the interview, they gave full and clear narratives all of which were based on their personal lived experience with Bob which I then coded into visual images that represent quality and which could be read by my new decoding system, The Eisner Image Reduction System.

After doing a varimax rotation to identify factors, the following factors emerged. They account for eighty two percent of the variance. Actually, between you and me I wanted to do a qualitative study but my university won't accept them, so I had to resort to numbers. Anyhow, I know you're dying to know the factors. They are:

1. Bob makes you feel a part of a family.
2. Bob stays in touch.
3. Bob is an iconoclast you can respect.
4. Bob finds a place for you when there is no more room at the Union.

I think these findings are clearly valid. They match exactly my own prejudices about Bob. Bob does make you feel a part of a family. He does stay in touch. He is a man whose convictions you can argue with. Bob does find a place for you when you need one—and so does Bernadine.

Who said you need a large random sample to get at the truth—not Bob, that's for sure!

And so my friends, my family, let me give you now an iconoclast, an outlier, a special case, someone you can argue with and win or lose come away having learned—that you lost something. Friends, Here's Bob!

Hoax?

Bob Stake[1]

Checking the long list of plans sometime last winter, Lizanne asked a title for my closing remarks. Having no idea, I pondered. She scribbled on, finally saying, "Tell me, Bob, of all the great people you have worked with in CIRCE, whose ideas influenced you most?" Now I had two ponderables. She went on writing. Finally, I said, "Maybe Hoke's." Turning to me, she said, "Hoax?" "Yeah, I guess, Hoke's?" "With a question mark?" "I suppose so."

From the bottom of my heart, I want to thank Lizanne for creating these two days, a marvelous deployment, and her staff, Connie, Karen, Trudy, Susan, Diane, and Beena, with great help from Elizabeth Easley, and hard work from Edith Cisneros, Marya Burke, Rita Davis and Terry Souchet. I appreciate the generosity of the Jack Easley Endowment, the Daniel A. Alpert Fund, and the Bureau of Educational Research. And I thank you all for coming, for speaking, for making it an honor for me, a delight for friends and family, a reunion for all.

To Teach. My mother and father were teachers. It only lasted a year for Grandpa Earl because some of the boys in his one-room schoolhouse were bigger than he. My mother taught for 15 years, the early ones in a sod schoolhouse in western Nebraska. Her grandfather had gone to the Genoa Indian Reservation in 1851 to bring agricultural methods to the Pawnee boys.

But I had no aspiration to teach. To pontificate, yes. To "show off," yes. But the thought did not occur to me until I needed a post-baccalaureate year to attain my Navy ROTC commission. I told Dean Henzlik I should use my non-Navy, available, upcoming, 28 credit hours to get a teaching certificate. He said, "Why?" I was stumped for an answer--

---

[1] This is my presentation to conclude a splendid symposium honoring my career at the time of my formal retirement on May 9, 1998.

but said, "I might get Navy duty with training responsibilities." He said, "That's an answer," and arranged it. So two semesters later, I was certified and commissioned at the same time, one day before I married Bernadine.

Bernadine soon was teaching in San Diego while I sailed Korean waters. Back in San Diego, I was impressed by one of my eldest cousins, Richard Madden, a professor of education at San Diego State and co-author of the Stanford Achievement Tests. Richard would spread his charts on a table in his study and show me test score trendlines of the children of Cherry Creek, Colorado, explaining how changes in the teaching of spelling had reconfigured the scores. I marveled at Richard finding connections between teaching and testing. Twenty years later, I hadn't found such connections myself, nor had my colleagues. For his dissertation research, I talked a bright, mature Aussie, Norman Bowman, into searching for present-day Richard Maddens, the practitioners so immersed in testing *and* curriculum that they could actually use the school's testing program diagnostically. Like Diogenes, he found none.

And ten years later, for her dissertation research, I persuaded a bright, mature Brazilian, Penha Tres, to study the interactive knowledge of testing and curriculum improvement at the Office of the Illinois Superintendent of Public Instruction, in Springfield, to find the people who would understand both assessment and teaching, so that tests would be built partly to serve a diagnostic purpose. And she found none. And although the efforts to build the IGAPs were harmonious with those of curriculum professors here at the University of Illinois and at other leading teacher training institutions, there was no study of consequential validity--so that it could be said with assurance that improvements in teaching will be manifest in changes in test scores.

Paraphrasing Milton: They also serve who leave the null hypothesis tenable. Just a few hours ago, Michael Scriven noted that it is a sophisticated researcher who beams with pride having, with thoroughness and diligence, found nothing there.

Understanding Testing. In 1954, my cousin would not let me enroll at San Diego State, saying there were better places to learn about testing. I was accepted for graduate school here at the U of I but, in a scorching August visit, somehow failing to meet Tom Hastings and Lee Cronbach, finding rent an unbelievable $125 a month, Bernadine and Jeff and I settled elsewhere.

A year later, a graduate assistant at the Educational Testing Service, I continued my fascination with test items. It was a while before I realized these items were just another version of showing off. I could devise analogy items that stumped even the cleverest of my friends.

As a political venture, I saw testing as "emancipatory." Poor youngsters who could solve analogy items could share in the affluences of society. It was another while before I realized that for every child enriched, many were further locked-out of privilege, lured by the winsome toils of analogy items.

Let me assure you that these tests had respectable validity in the sense that, for a large heterogeneous group of youngsters, the scores correlated well with subsequent grades in school. But as many of the critics of testing have noted, such test scores did not correlate well with success in later work, with practical ingenuity, aesthetic sensitivity, raising a family, being a good citizen, or becoming an effective teacher. And many of the people who became good at these other things found life harder because *their* test scores suggested *their* aspirations were less worthy of support.

My studies at Princeton concentrated not on test development but on psychometrics, mathematical theories of measurement of human characteristics. I wasn't very good at this stuff and it could be said that that was the reason I not only got out of testing, but became less reliant on quantitative measurement. Who knows? I returned to my alma mater, the University of Nebraska, to teach and do research. It is hard to believe these days, but Charles Neidt had held a tenure track position open for me for three years while I was getting a doctorate.

There at Nebraska, I did my research on instruction. I don't know why. I found it good to design highly structured, experimental, standardized studies of teaching. Somehow word got to Tom Hastings, whom I still had not met. Tom needed someone to succeed Dave Krathwohl and Phil Runkel as his assistant for the Illinois Statewide Testing Program, headquartered across the alley from Newman Hall. And what he wanted was someone who knew instruction and testing and might help make the Illinois tests more relevant to teaching. Who he needed was Richard Madden, but he got me. I arrived as he and Lee Cronbach were answering a US Office of Education invitation to create a National Educational Laboratory on campus, a CIRCE. Tom wanted it to emphasize connections between teaching and testing, Lee wanted it to emphasize connections between curriculum development and evaluation. I was so out of it that I doubt if a single paragraph I wrote got included in the proposal submitted by Lee, Tom, and Jack Easley.

One day as Tom and I were crossing a bike path on Wright Street, he asked me, "Now that you have learned to look both ways, what do you want to accomplish at Illinois?" I said, "I never think that way." It wasn't a premonition of going beyond goal-based evaluation. It was more like realization that success came easiest by setting low goals.

The Company. At CIRCE, Tom and I tried to help Mike Atkin, Bill Creswell and a number of national curriculum project leaders with their evaluation obligations. Jack somehow managed to get student responses analyzed and back in two weeks to Max Beberman's lesson writers, but that was still too slow. And time and again, the longer evaluations showed no significant differences. One answer was to do studies too small for inferential statistics. That may have been the origin of case studies.

Or it may have been the day Lee got out of the car at the Union, saying, "What this field needs is a good social anthropologist." It took me at least ten years to get an inkling of what he meant. But I didn't wait that long to pay attention to what Lou Smith and Barry MacDonald and Ulf Lundgren and Mariann Amarel were doing.

Early days at CIRCE were heady times. Jim Wardrop, Gene Glass and Doug Sjogren came aboard, then Ernie House, bringing Joe Steele, Tom Kerins, and Steve LaPan. Tom Maguire and Peter Taylor were first in a stream of splendid graduate students, Dennis Gooler and Mary Ann Bunda, Terry

Hopkins and Duncan McQuarrie. And so many more, Jennifer McCreadie, Oli Proppe, Jim Pearsol, Judy Dawson. And on and on.

Off and on for many years, Gordon Hoke and Terry Denny hung out with us; Claire Brown, Arden Grotelueschen, Jim Raths, Bob Linn. Bernadine headed a three-year evaluation of the National Center for Sex Equity Education in Fort Lauderdale. Jacquie Hill, Buddy Peshkin, Wayne Welch, Jim Sanders, Lou Smith, and Rob Walker helped with Case Studies in Science Education.

And a stream of head-turning visitors from far continents: Ulf Lundgren, Barry MacDonald, Peter Fensham, Helen Simons, Arieh Lewy, David Hamilton, John Nesbitt, Royce Sadler, Marli Andre, Don Hogben.

All of them, locals and aliens, wonderful teachers. From these, my personal mentors, I skimmed away over three thousand major ideas--acknowledging seven, six if you don't count Cronbach's curbside remark. The reason I said "Hoke's" was that over a thousand of the ideas were from Gordon alone, which he in turn had stolen, but he always included the citation.

I didn't learn how to teach in my semesters at Nebraska. I learned from you. And I learned from my mistakes, at which you didn't laugh. Well, Ernie did. But most of you just smiled and said, "That's real nice."

Metaevaluation. So I gradually learned that educational evaluation can't be done. It cannot be "done done." It's an impossible dream. If Ten is full-and-accurate determination of the value of an educational program, we sometimes get to Three, usually not past Two. The RFP calls for Michelangelo, and we are finger-painting. (I think I stole that line from you, Michael.)

We differ among ourselves as to the meaning of the words, "to evaluate," and we advise folks to do a lot of different things in the name of "evaluation." But speaking simply, it means to determine the quality of something. Everybody evaluates all the while: "You there are wearing your best shoes." "That melon at lunch tasted so good."

"Although fictitious, this morning's accolades were so gracefully put!" Or, the student in my 498 class this spring writing in her journal, "How can we learn to represent teaching quality when he won't tell us what it is?" Each of us is a constant producer of evaluations.

But professional evaluation, where we move well beyond common sense and impression, when we reject simplistic indicators; professional evaluation, where we propose to combine the discipline of the connoisseur, the logic of the philosopher, the acuity of the ethnographer, and the moral sensitivity of the judge. We are promising something we cannot do.

I look back over CIRCE's 34 years and wonder if we ever came close. We have spun some provocative webs. We have been temporarily familiar with a lot of teaching. We have fashioned some penetrating issues, told some good stories, written some handsome reports, occasionally been useful. But how close did we come to pinning down the merit and shortcoming of those programs?

I don't consider this an exercise in postmodern cynicism. Oh, I have my poststructuralist streak. Constructivism has its thrall, sometimes as tasty as ice cream. But I walk down the stairs a modernist. What I say today is, I believe, however deluded, a realistic metaevaluation of the field.

Analysis. I am not put off because we find a thousand notions of what good teaching is. Complex representations, we can handle. I am put off because we cannot agree that the whole is greatly different from the sum of its parts. And the embracing view of value is not nicely represented by a few exquisitely selected criteria. We are especially weak when we focus on but a few of the many parts.

For diving: the aggregate of perpendicular entry and small splash do not tell the quality of the dive. For creative writing: grammar, sequentiality, illustration, and closure do not tell the quality of the essay. And description and judgment of antecedents, transactions, and outcomes do not encompass the quality of the innovative instruction.

We cannot take solace in the fact that the most of the world doesn't want to know more about diving, or essays, or instruction. There is market for exit polling, for Dow Jones, for the vignette, for the sound bite, for simplistic representations. As Linda Mabry said this morning, all indicators are misrepresentations, but worse because they satisfy the curiosity for knowing the real meaning of the matter. Even the best of our evaluations allow people to falsely presume that a complete evaluation has been done.

We should not be satisfied that quality of teaching is known by student ratings, or by student test scores, or by peer reviews, or by teacher of the year awards. Teaching is a situationally responsive act, a role a hundred times more complicated than the best checklist or set of standards. Its meaning is constructed by the folks-involved every bit as much as the meaning of mathematics is constructed by children. The value is embedded in the situation, only in small part accessible to evaluators, supervisors, or the teachers themselves. Every child is shaped in part by teachers, for good or not, and most of the good they do, and most of the ill they do, is God's truth alone.

Does that mean it's been a waste? Of course not. We have done--not the best we might have--but many things worthy of pride. We know much more now that we did in the 60s. Thanks to Michael especially, and to many of you toilers here today, we have real help to offer program directors and constituents, help both toward the determination of value and the facilitation of self-study. And while preserving the connection Ralph Tyler made between the curriculum and evaluation, as Lee and others have said so persuasively these two days, we have brought democracy to the center of our conversation.

In-service. What did I learn? If I were to name the biggest thing I have learned in this time it is--it's what Ernie said this morning in different words--that the program and its value are one and the same, that *the meaning of an evaluand and its quality are one thing, not two.*

When I wrote the "countenance paper," I put description on one side and judgments on the other. But it was a mistake to imply that descriptive data and judgment data *should be* pulled apart. As we observe teaching, learning, the politics and the culture of education, we simultaneously see their merit and shortcoming. We can identify criteria and get ratings or scores, but these analytic calculations draw us away, I think, from understanding the quality of the program.

Our minds will analyze, analysis is a fixture, often useful to get us to attend to understudied parts, but analysis is construction as much as dissection. Values analyzed can be less a refinement, more a replacement. I continue to endorse "responsive evaluation" for its holistic mindset, responding to the activity, the complexity, the situationality and the quality of education with the fullest interpretation 180 pages will allow. But no one approach is good enough. As Oli Proppe said in his dissertation proposal, a dialectic among several mindsets is essential to good evaluation.

When we studied science education in the nation's schools in the 70s, we were up against a federal formula saying that quality is the difference between where we are and where we ought to be. But quality is not a discrepancy. It is an inherent, evolving, compounding of the evaluand.

As we have examined the quality of professional development at the Chicago Teachers Academy, we have found the merit of teaching and learning captured neither by Bill Bennett's "worst schools" soundbite, nor Bill Clinton's praise, nor Paul Vallas' probation list.

In education reports of all kinds, the executive summary is a fiction. Reality is at least "touched" by the description of teaching integrity and wasted opportunity in the classroom.

Beauty is in the eye of the beholder, but inseparable from the flower.

For me, it's been a great ride.

And Yet. Just as my analogy items made life harder for those who scored low, formal evaluation, as we have practiced it, has made Education less effective. If we put the power beams of consequential validity down on Evaluation, we see that we have failed to make it clear that almost everyone has too narrow a view of teaching and learning. And that narrowness distorts the judging of our youngsters, our schools, our society and ourselves.

At the top of the list of deceits we have failed to expose are those of standardized testing. We have failed to show that the best testing has regularly not been an indication of what students can do, nor of the quality of the educational system, nor of what the teachers or the society should do next.

According to Gallup Polls in the 60s, the populace had high confidence in our schools--now, grave doubts. In some ways, the schools are not as good as they were; in a few ways, they are better. But the image of the schools has changed, partly because the schools won't adapt to an evolving society, partly because many people don't want them to change as much as they do. An awful lot of people feel they know how to run the schools better. And a good part of the false confidence is at *our* doorstep. We most responsible for the formal evaluation of education have not provided better representation of teaching quality than standardized test scores.

Homework. So I ask your help. My colleagues are passing about the room handing out forms (see attachment here). Here is what we are going to do. We are going to do a study to help legitimate a fact that almost everyone in this room knows: *That you cannot use standardized student achievement test scores to determine quality of teaching.*

Each of you--should you accept the mission--will approach the principal of an elementary school, and, after pledging confidentiality and gaining rapport, ask him or her to identify one situation in which a quite good teacher has preceded or succeeded a teacher clearly not so good. That is, identify a classroom in which the teacher one year and the teacher the next year were of quite different teaching capability. Then you need to get the principal to release to you the test scores for that one room for the two years. With assurances that the assignment of students to that room has not changed, we can expect there to be a random plus and minus difference in means across the two years. Some of you will want to make several comparisons.

QTASTS Study

This is the beginning of a national study of professional development, quality of teaching and student performance in elementary schools, an effort to look at indicators of faculty involvement in continuing professional education. It will include an examination of the quality of elementary school teaching as represented in various ways including by student standardized test scores.

Sch:
Grade:
Content:

Professional Development > Classroom Teaching > Student Performance

Test:
Raw scores,
Scaled scores; G Eq

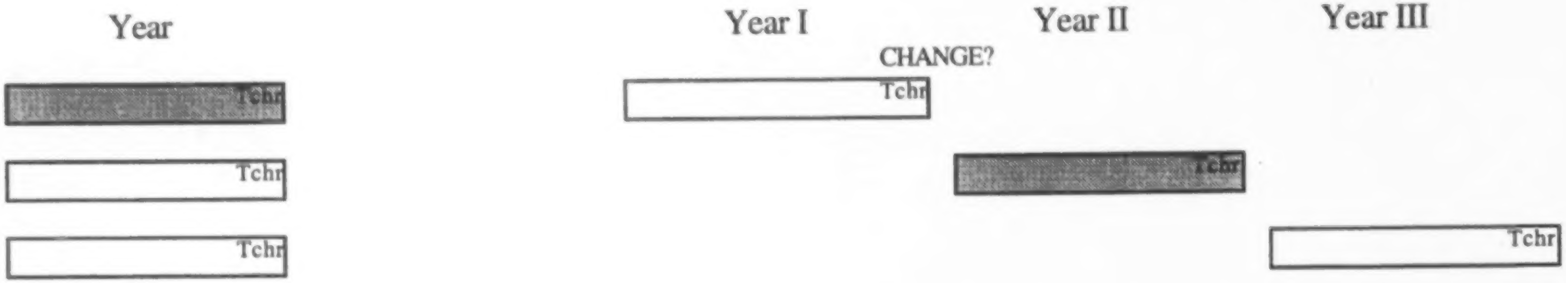
We intend to look particularly at what happens in the classroom after intensive involvement in professional development. We will also look at test scores when a participating teacher leaves and is succeeded by a non-participating teacher, or vice versa. Change in the other direction will be considered too. Still other comparisons will be done to relate teacher improvement and student performance.

Year I:	Year II:	Year III:	Year IV:	Year V:
IN THE SAME CLASSROOM		WITH A CHANGE	IN TEACHERS	
Teacher Participating - H M L 0 high quality - higher quality low quality - lower quality N = Mean/Mdn =	Same Teacher N = Mean/Mdn =	Same Teacher N = Mean/Mdn =	Next teacher N = Mean/Mdn =	That teacher Participating - H M L 0 high quality - higher quality low quality - lower quality N = Mean/Mdn =

Classroom scores from the shaded boxes are compared to those in other boxes. It is expected that usually the identification of teachers will be done by the school principal, sometimes augmented by views of other teachers, sometimes confirmed by observations of the researcher.

Eight characteristics of good teaching will be used to help specify areas contributing to high quality teaching.

> = <	Caring about kids	H A L
> = <	Creating Lng situations	H A L
> = <	Knowing subj matter	H A L
> = <	Classroom discipline	H A L
> = <	Teaching for the test	H A L
> = <	Getting good test scores	H A L
> = <	A good team player	H A L
> = <	Reflective, improving self	H A L



The study is to be done in non-departmentalized elementary schools where one teacher teaches the whole class at least 75% of the teaching periods. In the classrooms studied, children should be assigned more or less at random. The school needs to have administered the same standardized achievement tests for several years. Comparisons will not be good if there are ongoing changes in student demographics.

R. Stake, CIRCE, University of Illinois, 4/23/98

BEST COPY AVAILABLE

By finding no grounds for rejecting the null hypothesis, we will have a handsome citation that either student testing does not indicate teaching quality, or that principals do not know good teachers from bad. Obviously we will have to deal with several complications here, but it is time we made the citation.

I am serious. I have only a few research projects still to do, maybe one. Enough of vision; it is time for damage control. The aim is clear, to help improve popular conceptualization of school quality. I really would appreciate your help. This is no hoax.

Last word. We are not gathered here for commencement. Things *are* winding down for this teacher. The archivists will soon be by. They will look in my files and on my shelves, and find precious little to preserve. But it is not they who evaluate a career. What matters is in the eyes, the minds, and hearts of those I see before me today. In the words of Jennifer Greene yesterday, "Let's, you and I, 'toil on.'"

Thank you.